

AD-A147 553

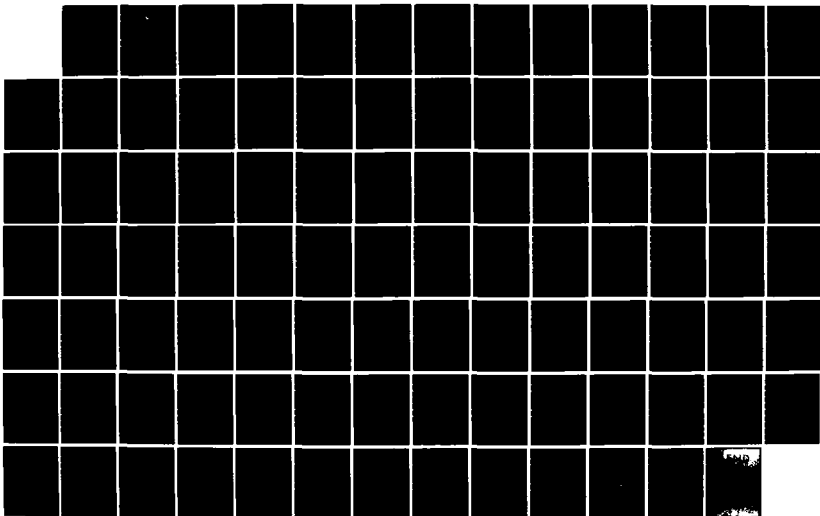
THE EFFECT OF ASSUMPTIONS ABOUT COST ELEMENT
PROBABILITY DISTRIBUTIONS ON. (U) AIR FORCE INST OF
TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYST.
D B FREEMAN SEP 84 AFIT/GLM/LSM/845-24

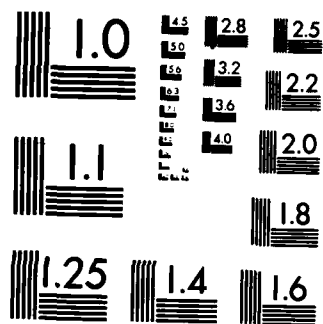
1/1

UNCLASSIFIED

F/G 15/5

NL





2



AD-A147 553

DTIC FILE COPY

THE EFFECT OF ASSUMPTIONS ABOUT
COST ELEMENT PROBABILITY DISTRIBUTIONS ON
OPERATING AND SUPPORT COST RISK ANALYSIS

THESIS

David B. Freeman
Major, USAF

AFIT/GLM/LSM/84S-24

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

DTIC
NOV 15 1984

Wright-Patterson Air Force Base, Ohio

111 14 11 11
REPRODUCED AT GOVERNMENT EXPENSE

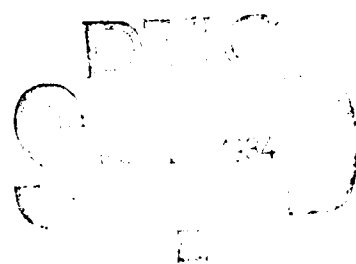
AFIT/GLM/LSM/84

THE EFFECT OF ASSUMPTIONS ABOUT
COST ELEMENT PROBABILITY DISTRIBUTIONS ON
OPERATING AND SUPPORT COST RISK ANALYSIS

THESIS

David B. Freeman
Major, USAF

AFIT/GLM/LSM/84S-24



Approved for public release; distribution unlimited

The contents of the document are technically accurate, and no sensitive items, detrimental ideas, or deleterious information are contained therein. Furthermore, the views expressed in the document are those of the author(s) and do not necessarily reflect the views of the School of Systems and Logistics, the Air University, the United States Air Force, or the Department of Defense.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
Distribution/	
Availability Codes	
Avail and/or	
Special	
A-1	



AFIT/GLM/LSM/84S-24

THE EFFECT OF ASSUMPTIONS ABOUT
COST ELEMENT PROBABILITY DISTRIBUTIONS ON
OPERATING AND SUPPORT COST RISK ANALYSIS

THESIS

Presented to the Faculty of the School of Systems and Logistics
of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Logistics Management

David B. Freeman, B.S.

Major, USAF

September 1984

Approved for public release; distribution unlimited

Preface

A dissertation by Lt Col John A. Long, coupled with discussions with Mr. Tom Recktenwalt of the Office of VAMOSC, inspired this effort. Risk analysis has become an integral part of the cost estimating process, but virtually all known methods of risk analysis rely on knowledge of the probability distributions of the cost elements. As yet, the distributions are unknown; therefore, the distributions must be assumed. Desmatics Inc., under contract to the Air Force, is attempting to determine those distributions, but there is an underlying problem. No one has determined whether or not the type of distribution will actually affect the risk analysis. In effect, then, the Desmatics effort may not be warranted. This study was conducted to determine the effect of different cost element distributions on risk analysis.

The experiment was performed with a Monte Carlo simulation. Three cost elements, fuel costs, maintenance personnel costs, and depot maintenance costs, were used as the sample space. Historical data from 41 fighter aircraft was used to determine the low, high, and modal values for each cost element. The data values established parameters for candidate cost element distributions. The simulation sampled the various cost element distributions and summed them. The result was an array of 18 total cost distributions containing 1000 data points each.

Analysis of the results indicated that the cost element distribution chosen does have an effect on the risk of the total cost estimate. Further, distributions with infinite upper bounds result in consistently higher risk than those with finite upper bounds.

I would like to thank Lt Col John A. Long for his assistance in this effort. His knowledge, encouragement, and patience with me were invaluable. Many thanks also to Roy Wood for 'filling the gap' by providing his services as a reader for this effort. In addition, my close friends warrant my appreciation, especially Wendy, who was always confident of my ability. Last, I would like to acknowledge my children, David and Katherine, for their sacrifices this past summer.

David B. Freeman

Table of Contents

	Page
Preface	ii
List of Figures	vi
List of Tables	vii
Abstract	viii
I. Background	1
Perspective	1
Life Cycle Cost	2
Concepts	2
VAMOSC II	8
The Role of Risk in LCC	9
Research Questions	11
II. Methodology	13
Fundamentals	13
Cost Equations	13
Cost Element Structure	16
Risk Measurement	17
Analysis Methods	19
Baseline Data	22
Experiment Overview	22
Data Point Selection	22
Cost Driver Determination	24
Constant Year Dollars	25
Candidate Distributions	26
Distribution Basis	26
Distribution Criteria	27
Beta Distribution	28
Rectangular Distribution	30
Triangular Distribution	31
Gamma Distribution	32
Selection and Validation	33
Monte Carlo Simulation	35
Justification	35
Program Structure	36

	Page
III. Results and Findings	40
Validation Results	40
Simulation Results	45
Findings	47
Research Question 1	47
Research Question 2	48
Research Question 3	49
Follow-On Research	49
Summary	51
Appendix A: Data and Constant Year Dollar Factors	52
Appendix B: Cost Element Parameterization	55
Appendix C: Monte Carlo Simulation Source Code	58
Appendix D: Triangular Distribution Calculations	62
Appendix E: Fuel Cost Element Histograms	64
Appendix F: Maintenance Cost Element Histograms	68
Appendix G: Depot Cost Element Histograms	71
Appendix H: Kolmogorov Test Source Code	75
Bibliography	78
Vita	81

List of Tables

Table		Page
I.	Sampling Combinations	37
II.	Comparison of Means and Modes	42
III.	Comparison of Variances and Ranges	42
IV.	Kolmogorov Test Results	44
V.	Rank-Ordered Variances	46
VI.	Gamma Distribution Effects	47

Abstract

The purpose of this study was to determine the effects of assumptions about cost element probability distributions on Operating and Support cost estimates. Currently, these assumptions are being made arbitrarily, without regard for their effect on the risk associated with a total cost estimate.

The experiment was performed with a Monte Carlo simulation. Three cost elements, (fuel costs, maintenance personnel costs, and depot maintenance costs) were used as the sample space. Historical data from 41 fighter aircraft was used to determine the low, high, and modal values for each cost element. The Triangular, Beta and Gamma distributions were selected as candidate distributions for the cost elements. The low, modal and high values provided a means by which to parameterize the distributions for specific cost elements. The distribution for each cost element was varied over the set of candidates with all other factors held constant, samples were drawn and then summed to provide total cost estimates. The result was an array of 18 total cost distributions containing 1000 data points each.

The variance of the 18 resultant distributions was the focus of the analysis. That analysis indicated that the cost element distribution chosen does have an effect on the risk of the total cost estimate. Further, distributions with infinite upper bounds result in consistently higher risk than those with finite upper bounds. ↗

THE EFFECT OF ASSUMPTIONS ABOUT
COST ELEMENT PROBABILITY DISTRIBUTIONS ON
OPERATING AND SUPPORT COST RISK ANALYSIS

I. Background

Perspective

Some of the most significant decisions being made in the Department of Defense (DoD) today concern the acquisition of new weapon systems. When the decision is made to incorporate a new system in the defense posture of the United States, the government incurs not only the cost of acquisition, but also the cost of operating and maintaining that system for its operational lifetime. These decisions are based, at least in part, on cost estimates. Since these estimates are attempts to predict future costs, there is a degree of risk associated with them.

During recent years interest in the quality of decisions made, both in the public and private sectors of the economy, has grown tremendously...In particular it is clear that traditional approaches to decision making are lacking in certain dimensions, particularly in the manner in which they commonly fail to link together initial and consequential alternatives, the way that uncertainties are dealt with on an informal basis and the evaluation of information in an arbitrary method [32:ix].

Public and Congressional scrutiny of DoD decisions is easily justified. The proposed Fiscal Year 1984 Federal Budget included a DoD request for \$124.5 billion for systems procurement and an additional \$74 billion for operation and maintenance of systems. Together, these figures comprise about 75% of the total defense budget and approximately 20% of the total

Federal Budget (17:8; 26:41,76). The financial ramifications of a poor decision in DoD are not easily dismissed because of the long term commitment involved with the systems.

Traditionally, lowest production cost, without regard to total system cost, has been the financial criterion used in system selection. The fallacy in this approach is that it ignores the outyear expense of the system. In fact, it can imply a greater outyear cost because of less reliable systems due to cheaper parts and materials. Only recently has total cost been used in the decision process. However, total cost estimates have not been well received, in part because of the large amount of uncertainty associated with them (2; 8).

Methods do exist to treat the uncertainty, but their use has been limited.

No method, short of eliminating uncertainty, will guarantee success in every situation, but an increase in the percentage of correct decisions made can have a dramatic effect on an organization's overall results [32:ix].

An understanding of these methods requires an understanding of both the content of total system cost and the origin of the uncertainty in estimating that cost. For the purposes of this effort, discussion will focus on the United States Air Force; however, the concepts apply to the entire Department of Defense as well as to the private sector.

Life Cycle Cost

Concepts. Life Cycle Cost (LCC), as an imperative in the Air Force acquisition process, resulted from high cost overruns of major systems acquisition in the late sixties and early seventies (27:30). In general

terms, "LCC is the search for the significant costs that can be influenced by planning and design decisions" (37:18). More specifically, Air Force Regulation 800-11 defines LCC as

...the sum total of the direct, indirect, recurring, nonrecurring, and other related costs incurred, or estimated to be incurred, in the design, development, production, operation, maintenance and support of a major system over its anticipated useful life span [12:31].

Initially, performance specifications are the only precise information available about a proposed weapon system. It stands to reason that a LCC estimate based on limited information will be commensurately uncertain. As the system progresses through the Concept and Validation phases of the Air Force acquisition process, more and more detailed information is developed. The LCC estimates gain precision in parallel with the amount of information known (37:9,15). It also stands to reason that a method used to compute LCC estimates early in a system's development would not be the same as a method used later. There are, in fact, three common methods of LCC estimating.

In the early stages of development, analogy is the most common method of estimating LCC (38:16). In this case, costs of similar, previously fielded systems are used to infer the cost of the new system. Problems arise, though, when the new system will incorporate "leading edge" technology or new materials (37:15). When no analogous or explicit cost information is available, "ball park" figures are used; thus the precision of the LCC estimate in this case is limited (9:3-3).

The second method of estimating LCC is known as the parametric method. It makes use of various mathematical processes to develop Cost Estimating Relationships (CERs). One method of developing CERs is

regression which yields equations relating specific program characteristics and cost categories to explanatory variables (3:372-373).

If there are prior hardware systems which can be compared with the new (proposed) system, and if physical, performance, and cost data are available on the older systems, then statistical analysis may provide useful cost projections [5:3-3].

These CERs are usually specific to a single type of system. For example, in the aerospace industry, it is known that aircraft cost can be related to weight, thrust, and speed (37:50). CERs may also be developed using a common sense approach. For example, logically, a simplistic personnel cost could be determined by summing the wages of all the workers. The CER might be simplified by taking the average number of workers times the average wage. The primary weakness of the parametric method is that it requires previously collected, exhaustive data (38:16).

The final method of LCC, the most detailed of the three, is item costing or the Engineering Design method. Once detailed specifications are developed by engineering teams, costs can be estimated with a degree of accuracy not available in the other methods (37:16). The cost estimates are developed starting at the component level and working up to the total system. This is known as a 'bottom up' approach. The strength of this method lies in the fact that it can be used to evaluate Engineering Change Proposals (ECP's) at almost any stage in the acquisition process where detailed specifications are available (5:47). However, it must be noted that this method usually cannot be used early in the acquisition process when detailed design information is commonly not available. Additionally, the focus of this method is the actual

system and sub-systems, consequently, it may ignore important peripheral cost considerations (e.g. support personnel costs) (38:16).

Modelling is a tool that has been developed to aggregate the costs determined by the three methods above. The models allow variation of the components providing sensitivity analysis of the cost estimate.

Four common types are used in the acquisition process.

1. Optimization -- These models are used to analyze specific de-parameters or cost factors. Their drawback, in the context of total system cost, is that they look only at a microcosm and not at the system as a whole (38:16,21).
2. Parametric -- These models use CER's to predict total cost. They are relatively easy to use and inexpensive; however, CER's relating performance to operating and support costs are not available (38:19).
3. Simulation -- Actual operational situations are recreated on the computer to provide a set of statistics on any of the issues being modelled. Although of great value in sensitivity analysis, the idea of quantifying reality is somewhat suspect (38:16,21).
4. Accounting -- A set of equations to aggregate components is used to predict a single life cycle or operations and support cost. The most significant shortfall is that the equations have not been validated with real world costs (38:16,20).

To this point, the focus has been on how to develop a cost estimate. The value of the estimate, though, lies in how it is used.

According to Seldon (37:11), analysis of LCC can be used for:

1. Long-range planning
2. Comparison of programs
3. Comparison of support plans
4. Source selection
5. Program control
6. Trade-off decisions

OMB Circular A-109 stresses the importance of all six of these uses in

DoD acquisitions (34:4,5,10). In addition, DoD Instruction 5000.2 (10) emphasizes the importance of LCC considerations throughout the acquisition process by requiring estimates or summaries of LCC in the documentation for each phase of system acquisition.

Recently, General Robert T. Marsh, Commander of Air Force Systems Command (AFSC), summed up the emphasis within the Air Force acquisition process:

We must strive to procure our weapon systems in the most cost-efficient manner, and we must find means to arrest the ubiquitous cost growth of our weapon systems [31:1].

The inception of AFSC "Project Cost" in September of 1982 reflects that emphasis. This particular project has three main goals:

1. Affordability - to use every tool available to provide objective assessment of desired capability and the cost of that capability.
2. Stability - to minimize changes by developing a baseline of need and limiting subsequent changes to the system.
3. Management - to use cost control as a prime factor in the acquisition process and to attempt to resolve tradeoffs without cost growth. (31:1)

To lend further support to this philosophy, AFR 800-11 states that the Air Force should satisfy its needs using the system with the lowest LCC. Until recently, this was not the major thrust of concerns about system cost. Concentration had been on control of the acquisition cost, and analysis of long term operations and support (O & S) cost was very limited (16:36). This lack of focus on LCC resulted from the visibility of "up-front" or acquisition cost, combined with the relative invisibility of O & S cost. Often, design specifications or changes which would result in lower O & S costs (and thus lower LCC) were not considered viable because they required higher expenditures initially.

The visibility of those higher costs precluded the use of such options. In addition, there has been an absence of criteria with which to make such a tradeoff decision. Traditionally, program managers required a very high payback of funds expended "up front" to institute a change based on LCC (4:12,15).

The concept of designing a system to LCC, in many ways, ameliorates the problems associated with the visibility of acquisition costs. It takes the emphasis off production cost and forces all involved parties to consider cost throughout the acquisition process (16:38). The contractor is thoroughly involved because the Request for Proposal (RFP) contains provisions for using LCC as one of several criteria for source selection. In addition, awards can be determined using LCC estimates as baselines (25:30). Unfortunately, the first two attempts to use LCC as a selection criterion for a major system failed. Both the F-16 and A-10 programs required that LCC be a criterion, but a combination of poor/unavailable data, and incomplete information given to the contractors precluded use of the estimates in source selection (2; 8). The scarcity of available data is particularly important when trying to estimate the O & S component of LCC. A reasonable prediction of O & S costs requires a real world cost database with costs tied to the specific weapon system(s) that generated them (38:14). The Air Force does not currently have a large database of this type.

And it is a problem of fundamental importance, because a substantive cost analysis capability cannot exist without an appropriate information and data bank [19:24].

The dearth of applicable cost data is not a result of having no data, it is a result of having the right data in the wrong format. The Air Force

has a multitude of cost collection databases designed for functional area management, but the data is not matched to specific weapon systems (19:25-26). One solution to this problem is the major overhaul of the databases that exist.

VAMOS II. Visibility And Management Of Operating and Support Costs (VAMOS II) is the DoD solution to the data problem. In 1975, realizing that cost data collection and a standard cost element structure to support it was critical to future weapons acquisition decisions, DoD issued Management By Objective (MBO) 9-2. It specifically called for all services to make an effort to identify O & S costs by system. The intent was that if costs could be identified, then they could be controlled. In 1976 the Air Force developed the O & S Cost Estimating Reference (OSKER), which produced data annually in categories outlined by the Cost Analysis Improvement Group (CAIG) (36:1,6). CAIG's delineation of cost categories largely solved the problem of non-standard cost element structures (30:12). Within two years, serious deficiencies were identified in OSKER. As a result, VAMOS was developed. The primary objective of VAMOS was to collect and display costs, stated at the mission design series (MDS) level (for aircraft), in CAIG approved elements (36:3). VAMOS is not a new data collection system per se. It takes costs collected by functional areas and applies algorithms to split out and apportion costs by weapon system (36:3).

The efforts of VAMOS will eventually provide the database necessary for adequate O & S cost estimation; however, it will not eliminate the uncertainty associated with them. That uncertainty must be investigated, and it is the focus of the remainder of this project.

The Role of Risk in LCC

A decision is a choice among alternative courses of action. The decision-maker must attempt to predict the future and choose the action which provides the desired results (20:4). If future outcomes are known, the decision is made under certainty. However, far more often, decisions are made without knowing precise outcomes. Uncertainty, then, is the likelihood that the actual outcome will diverge from the anticipated outcome (21:9). Risk is the degree or amount of uncertainty in a decision. Realistically, the decision-maker will have some information about the future. Even if this information is subjective, it provides a range of outcomes. This range can be translated to a probability distribution, and from that distribution, risk can be quantified. This quantification is risk analysis (21:10; 22:324).

One of the primary uses of cost estimating is comparison of alternative choices (37:11). Too often, though, a decision-maker is presented with cost figures without any quantification of the risk. It is imperative that a decision-maker be appraised of the risk associated with LCC or O & S cost estimates; otherwise, he could easily make a bad decision. Take for example, estimates of O & S cost for two competing aircraft designs:

	<u>Aircraft A</u>	<u>Aircraft B</u>
Yearly O & S cost	\$9 million	\$9.9 million

Given just this information, a decision-maker looking for the most economical alternative would select aircraft A with a savings of \$9

million over a ten year system life ($.9 \times 10$). But suppose that the information were presented as a range of cost as follows:

	<u>Aircraft A</u>	<u>Aircraft B</u>
Yearly O & S cost	\$8.5-16 million	\$9.8-10 million

Now the decision is somewhat more complicated. Although aircraft A has a lower minimum cost estimate, it also has a significantly higher maximum cost (\$16 million). And although B has a higher minimum cost estimate, the most it could cost is \$10 million per year. Aircraft A represents a system with far more risk than aircraft B. Which system is chosen then depends upon how much risk the decision maker feels is justified or "acceptable". Judgement, after all, is the basis of decisions: analysis exists only to sharpen judgement (20:4).

The source of risk in the previous example is unknown. In order to quantify the risk, a deeper investigation of an O & S cost estimate is necessary. One major element of O & S cost for an aircraft is fuel. That single element might be computed as follows:

$$FC = HRS \times FCH$$

where

FC = yearly fuel cost
HRS = flying hours per year
FCH = fuel cost per flying hour

In like manner, all the cost elements could be computed and a total cost estimate would result. It must be stressed, however, that these apparently precise estimates are subject to considerable variability. In the example above, fuel cost per flying hour can only be estimated

because of the economic factors involved in pricing fuel, and because each engine has a slightly different rate of consumption. This implies that fuel cost is a random variable. In fact, each cost element is a random variable and thus, total O & S cost is a random variable also. As such it has a probability distribution that defines a range of values and the probability of each value that the cost can assume. For example, total cost may be estimated at \$14 million per year, but it could range from a low of \$8 million to a high of \$17 million. If the probability distribution of the total cost is known, this range represents risk for the decision maker. If the distribution is unknown, the range represents uncertainty (20:6).

The explanation above is simplistic and ignores a major problem: the distributions of the cost elements are unknown, and consequently, the distribution of the total cost is unknown.

A fundamental problem in life-cycle costing is the amount of uncertainty inherent in the estimation process as a function of the uncertainty of the inputs [18:9].

This presents a problem for the analyst who wishes to conduct risk analysis of an O & S cost estimate. The analyst must make an assumption as to the probability distribution of each cost element in order to conduct the risk analysis.

Research Questions

Any assumption made in an analysis, quantitative or qualitative, may introduce bias. In the case of O & S cost risk analysis, the effect of assumptions about cost element probability distributions is unknown. If the chosen distribution has an identifiable effect on the risk

analysis then this information should be considered. For example, if a cost element is assumed to have a gamma distribution and this assumption causes an upward shift of the total cost, then the decision maker should be advised that the stated risks may be biased. In order to quantify the effects of these assumptions the following research questions will be investigated:

1. Do assumptions about the probability distributions of cost elements affect the risk of an O & S cost estimate?
2. If a given assumption does affect the risk, is the effect consistent for all cost elements?
3. Can actual cost element distributions be determined with data presently available?

The remainder of this effort is directed at finding the answers to these questions. Chapter II consists of a detailed explanation of the methodology used to answer the research questions, and Chapter III will detail the results and findings of the experiment.

II. Methodology

The focus of this effort is the risk analysis of an O & S cost estimate, with risk defined as the amount or degree of uncertainty of the estimate.

A decisionmaker can be expected to want to know how much uncertainty is associated with the cost estimates he receives and how sensitive they are to changes in other variables that are themselves uncertain [30:14].

There are several different sources of uncertainty in cost estimates. These can be broken into two major categories: 1) uncertainty from variations in system design specifications, and 2) uncertainty from analyst bias, database errors, CER faults, and extrapolation (14:1-2). A cost analyst has little or no influence over the first category which comprises about 80% of total uncertainty (14:2). It is the second category which is relevant to this discussion. One of the prime contributors to this category is analyst bias in the form of making unrealistic or invalid assumptions (3:155).

The remainder of this chapter outlines the strategy employed to quantify the effect of analyst assumptions in terms of risk analysis. Some fundamental concepts will be discussed, and then the actual procedures used will be described in the order in which they were performed. Situations which required that assumptions be made will be discussed in the order they arose during the experiment.

Fundamentals

Cost Equations. Cost estimates can be derived from known factors

and rates, CERs (via analogy or parametrics), or expert opinion. The relationships, if used, can be either linear or non-linear. This experiment was restricted to the use of linear relationships. The general form of a linear relationship is

$$\text{Cost} = ax + b$$

where
a and b are constants
x is a random variable

Since x is a random variable, it follows that 'Cost' is also a random variable. A somewhat more complex form, known as a multivariate, is

$$\text{Cost} = X_1 + X_2 + \dots + X_n$$

where
 $X_1 \dots X_n$ are specific cost elements.

Each of the cost elements represents a specific category of cost such as fuel, personnel, or training for a given period of time (usually a year). Some or all of them could have coefficients (3:374). The summation of costs accrued is the basis of the Accounting method of modelling O & S cost estimates referred to in Chapter I (reference page 5). It must be noted that these cost elements are not constants. They are random variables whose values follow a certain probability distribution. If historic data is available for each of the elements, then a cost analyst can make an assumption concerning that distribution. This is precisely the technique used in this experiment.

Each cost element may be made up of one or more factors. The factors, as well as the method of aggregating them, are critical to risk

analysis. For example, a hypothetical cost for training three people might be expressed as follows:

$$\text{Cost} = 3(\text{TC})$$

or

$$\text{Cost} = \text{TC} + \text{TC} + \text{TC}$$

where

TC is the average training cost per person

Although this seems to be an insignificant difference, it has a great deal of relevance in risk analysis. For the first case:

$$\begin{aligned} E[\text{Cost}] &= E[3(\text{TC})] \\ &= 3(E[\text{TC}]) \end{aligned}$$

And for the second case:

$$\begin{aligned} E[\text{Cost}] &= E[\text{TC}] + E[\text{TC}] + E[\text{TC}] \\ &= 3 E[\text{TC}] \end{aligned}$$

However, the variances of the two equations differ. For the first case:

$$\begin{aligned} \text{Var}[\text{Cost}] &= \text{Var}[3(\text{TC})] \\ &= 3^2 \text{Var}[\text{TC}] \\ &= 9\text{Var}[\text{TC}] \end{aligned} \tag{1}$$

But for the second case:

$$\begin{aligned} \text{Var}[\text{Cost}] &= \text{Var}[\text{TC} + \text{TC} + \text{TC}] \\ &= 3\text{Var}[\text{TC}] \end{aligned}$$

The first cost estimate will have a variance nine times that of the second. Since risk analysis focuses on the variability of total cost, the method of collecting cost is, therefore, critical. It is pertinent to note that the cost database used in this study uses the additive method of cost accrual (36:6).

Unit Mission Personnel	Installation Support Personnel
Aircrew	Base Operating Support
Military	Military
Maintenance	Civilian
Military	Real Property Maintenance
Civilian	Military
Other Unit Personnel	Civilian
Military	Medical
Civilian	Military
	Civilian
Unit Level Consumption	Indirect Personnel Support
POL	Misc. O & M
Maintenance Materiel	Medical O & M
Training Ordnance	Permanent Change of Station
	Additional Duty Pay
Depot Level Maintenance	Depot Non-Maintenance
Airframe Rework	General Depot Support
Engine Rework	Second Dest. Transportation
Component Repair	
Support Equipment	Personnel Acquisition and Trng
Software	Acquisition
Modifications	Individual Training
Support	
Contracted Unit Level	
Sustaining Investment	
Replenishment Spares	
Replacement Support Equip.	
Modification Kits	

Figure 1. CAIG Cost Element Structure (1:8-16)

Cost Element Structure. Once costs for each element have been accrued, there is yet another factor which impacts the total cost estimate. That is, which cost elements to include in the estimate, and what comprises each cost element. This breakdown is known as a Cost Element Structure (CES).

In the mid 1970s, lack of a standard CES was identified as a prime deficiency of Air Force cost data collection and analysis (30:12). In response, the Cost Analysis Improvement Group (CAIG) published a CES for aircraft O & S costs as shown in Figure 1.

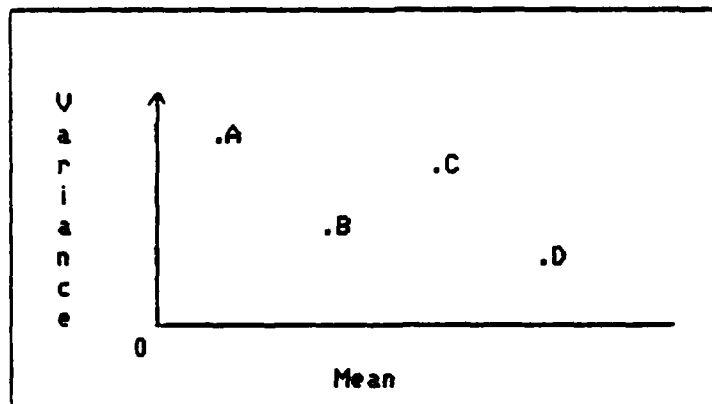


Figure 2. Mean - Variance Plot

Although all of these costs are important in an analysis, greatest effort should be directed toward those costs that are most affected by program decisions, distinguish two alternatives, and/or account for a substantial proportion of the total cost. The latter are known as cost drivers (1:7). The three most common aircraft system cost drivers are: 1) Unit Mission Personnel, 2) Aviation Petroleum, Oil and Lubricants (POL), and 3) Depot Maintenance (1:7).

Risk Measurement. Once the CES for an estimate has been determined (or selected cost drivers chosen), it is then possible to measure the risk associated with that estimate. Typically, the variance of the distribution of O & S cost for a system is used as a measure of the risk associated with the cost estimate for that system (21:10). Although variance can be depicted in a number of ways, there are three common types of display: a mean - variance plot, a cumulative distribution function plot, and a floating bar chart (29:108-113).

Figure 2 is an example of a mean-variance plot for four weapon systems, A, B, C, and D.

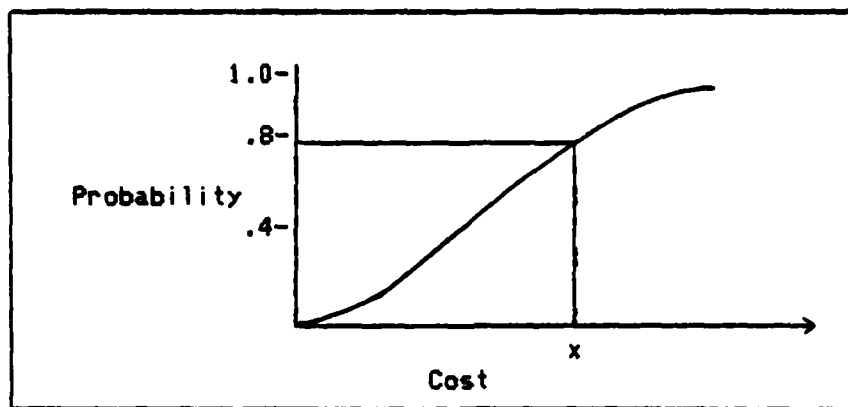


Figure 3. Cumulative Distribution Function Plot

A decision-maker can readily identify systems whose mean cost estimate falls within the range he desires. Then each estimate's risk can be assessed using the variance depicted on the vertical scale.

In Figure 2, system A has the lowest mean estimate but also the highest variance, signifying a relatively high degree of risk. System D has a higher mean, but has a low variance and subsequent risk. There is no clearcut 'best' choice, but the decision-maker can choose that system which best fits his criteria.

In contrast, a cumulative distribution function can be used to depict a single cost distribution's risk (variance). As shown in Figure 3, the decision-maker can determine the probability of the actual cost falling below the cost estimate using the vertical axis. This would be particularly useful if a system had already been chosen and the decision-maker was trying to determine a 'reasonable' cost estimate. Again, the graph merely depicts the data. The choice of an estimate depends on the decision-maker's criteria.

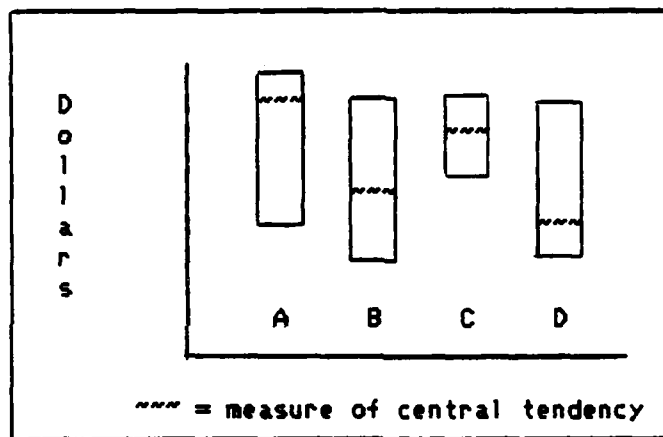


Figure 4. Floating Bar Graph

Figure 4 is an example of a floating bar graph. This combines the first two methods, and, consequently, it may be the most effective means of visualizing risk when multiple systems are involved. A portion of each system's cost distribution is displayed along with its mean (or other measure of central tendency). The decision-maker can easily weigh the differences of each distribution and select the one most suitable to his purposes.

Analysis Methods. In order to display information regarding risk analysis, the cost distribution(s) must be determined. Ideally, a large amount of historical data could be used to statistically determine the actual distributions; however, rarely is such data available. When the data is not available, the distributions must be determined in some other fashion. Insufficient data exists to unquestionably define an O & S cost distribution. This problem is central to this effort. There are, however, two methods of obtaining cost distributions on which risk analysis can be conducted.

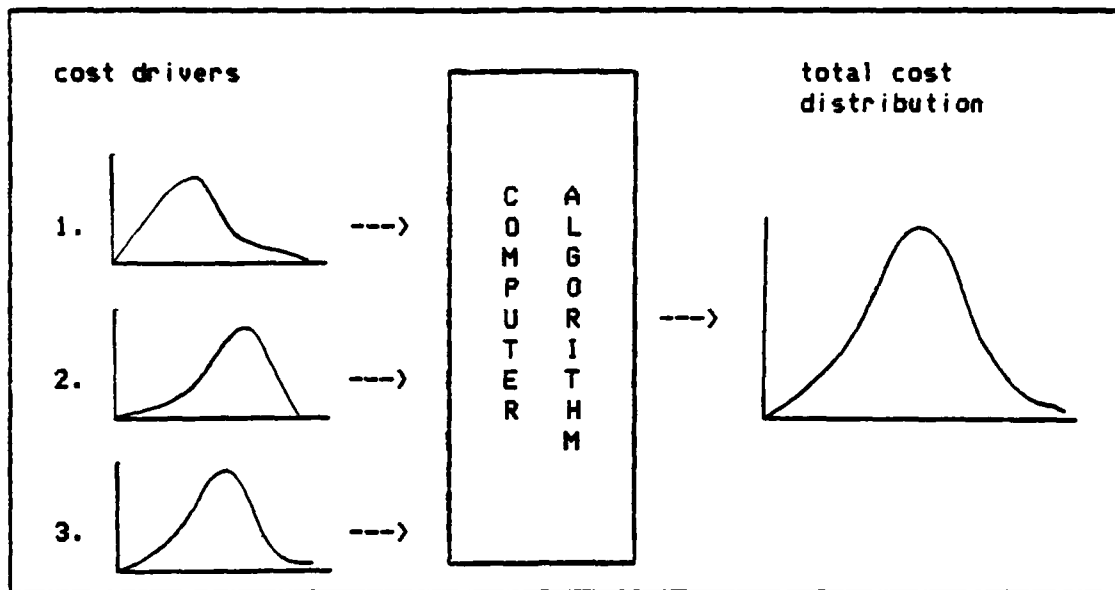


Figure 5. Monte Carlo Model

. LaPlace and Mellin transforms can be used to obtain an exact probability density function (pdf) for O & S cost. Given a set of cost factors, which sum to total cost, an analyst can derive the pdf for total cost. The result is precise but requires complex mathematical manipulation of cost factor transforms. In fact, this method can quite easily exceed the analyst's ability. For a more in depth review of this method, the reader is referred to Long (29).

A more functional, but less precise, method involves Monte Carlo Simulation. The key to the technique is the expression of cost elements/cost drivers as probability distributions around a mean value. The distributions must be assumed. The effect of this assumption is the main research question in this thesis. This technique, therefore, was an ideal medium for this experiment. Monte Carlo Simulation is an Input - Process - Output model.

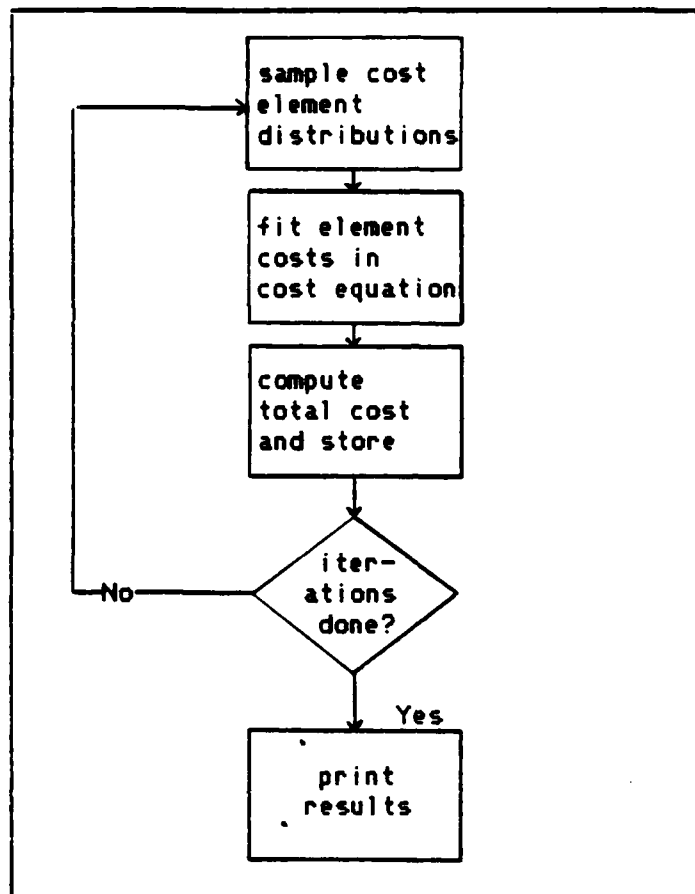


Figure 6. Monte Carlo Algorithm (14:17)

The cost drivers (and their respective distributions) are the input, a computer algorithm for sampling the distributions is the process, and the frequency distribution of the total cost is the output (see Figure 5) (14:6).

A major assumption of this technique is that the input parameters are independent of one another. For example, the analyst would have to assume that manpower levels were not affected by training costs. If dependence is suspected, it can be dealt with by incorporating the dependency relationship in the cost equation. Sampling from a joint

frequency distribution would also solve the dependency problem (14:11). Figure 6 is a simplistic computer algorithm for use in a Monte Carlo Model.

Baseline Data

Experiment Overview. With the fundamentals just described in mind, an experiment was designed to determine the effects of assumptions about cost element distributions on risk analysis of an O & S cost estimate. Data on O & S costs was collected from the VAMOS II HAF-LEY(A&AR)8203(DD) reports for fiscal years 1981 through 1983. The data was presented in CAIG format (reference page 16). Specific aircraft systems were selected for the experiment, and the cost drivers for those systems were determined. The data was then converted to constant 1983 dollars. Distributions were selected via commonly used statistical methods and a Monte Carlo simulation was developed to produce several cost estimate distributions. These resultant distributions were analyzed to find the answers to the questions posed in Chapter I. In the sections that follow, each specific step in the course of the experiment will be explained in detail and in the order in which it was performed.

Data Point Selection. The VAMOS II database contains O & S cost data for every aircraft in the active Air Force inventory. However, cost estimation is usually limited to a single aircraft category. The fighter category was selected for the experiment. Since the costs were to be reduced to dollars per flying hour so that cost element distributions could be chosen, the field had to be narrowed even more. It would be inconsistent to include the maintenance cost of a simplex

machine in a distribution with the costs of highly complex machines. The data would be skewed.

Two main criteria were used to achieve the necessary consistency in the data. The aircraft had to be dual engine and had to have approximately the same type of onboard systems. The dual engine criterion eliminated bias in the data caused by doubling single engine cost per hour. As an illustration, assume that an F-16A had a fuel cost per flying hour of \$750. If that was included in the baseline data with dual engine fighters whose costs ranged upward of \$1400 per hour, the resultant cost distribution would be skewed toward the lower costs. Although doubling the single engine cost might seem appropriate, there is no evidence to indicate that the assumption of linearity for fuel consumption is valid. This criterion eliminated the F-106 and F-16 aircraft.

The systems onboard the selected aircraft also had to be consistent. In this case, a simplex aircraft would probably experience a lower depot maintenance cost (again skewing the data). The aircraft selected had to have onboard weapon systems of similar complexity. Admittedly, this is a subjective assessment, but it was done on the basis of field experience. The AT-38 / F-5 series of aircraft, as well as the A-10, failed to meet this criterion.

The aircraft selected for inclusion in the baseline data are listed in Appendix A along with the raw data extracted from the VAMOS II database. In general, the F-4, F-15, and F-111 series of aircraft were included. The aircraft are all dual engine and have complex onboard systems. Further, their operating environment is similar even though

their employment profiles vary from interdiction to air-to-air combat to air-to-ground roles. Since all the aircraft operate in an arena not conducive to optimum engine operation, even more consistency was injected in the data. At this point one assumption had to be made: that the aircraft selected were mature systems (1:6). An immature aircraft might exhibit erratic or excessively high costs due to maintenance learning curves.

Cost Driver Determination. In the discussion on Cost Element Structure (reference pages 16-17) three elements were introduced as the most common aircraft cost drivers. They were Unit Mission Personnel, POL, and Depot Maintenance (1:7). Without exception, these three cost elements were the major contributors to the O & S cost of the aircraft selected. Consequently, they were chosen as the cost drivers used for this experiment. However, because of some inconsistencies in the missions of the aircraft, only certain portions of these cost elements were used.

Depot Level Maintenance was the only one of the three cost elements used in its entirety. Figure 1 (reference page 16) depicts the Cost Element Structure and all the subcategories. Unit Mission Personnel was restricted to military maintenance personnel. The aircrew costs were eliminated because some of the fighters selected were dual seat and others were single seat. The Other Unit Personnel category was also eliminated, but for inconsistencies in the data. Some aircraft had been allocated a cost due to Security Personnel but others were not. No reasonable explanation for this could be found; therefore, it was treated as an anomaly in the collection of data and was eliminated.

There were also inconsistencies in the Unit Level Consumption cost element. The use of training ordnance is restricted for the F-15 series of aircraft because of their air-to-air role whereas, F-111 and F-4 series aircraft routinely expend practice bombs. To enhance consistency in the data, only the PDL portion of this element was used. In summary, the three cost drivers selected for the experiment were PDL, Military Maintenance Personnel, and Depot Level Maintenance. Based on the data extracted from VAMOS II, these represented (on the average) 23.8%, 16.3%, and 17.8% of the O & S cost for the systems selected.

Constant Year Dollars. For a cost estimate to be meaningful, it must be expressed in terms the decision-maker can understand. A simple statement that a system will cost \$15 million over its lifetime means little unless the estimator specifies the cost as constant year dollars or then-year (inflated) dollars. If the estimate is in inflated dollars, that implies that the analyst has made an assumption about the rate of inflation over the life of the system. For this experiment, all costs extracted from the database were converted to 1983 dollars. Factors taken from AFR 173-13, 1 February 1984 (11:92), Table 5-1 were used to adjust the data. These factors are based on historic inflation rates and are specified in general categories such as fuel or military compensation. The general equation used to convert the data was

$$\text{FY83 Constant Dollars} = (\text{FY81 Dollars}) / (\text{Factor})$$

(The equation applies to 1982 data as well).

The conversion to constant year dollars was necessary for one reason: to eliminate the variability of actual prices for fuel, compensation, etc. from the cost equation. This could also have been

done by dividing the costs in the database by the item dollar price.

For example, POL cost can be depicted as

$$\text{Cost} = (\text{Flying Hours})(\text{Gallons/Hour})(\text{Dollars/Gallon})$$

If this is divided by the item price (Dollars/Gallon) then

$$\text{Cost} = (\text{Flying Hours})(\text{Gallons/Hour})$$

The resultant cost would be expressed in gallons as opposed to dollars. Although this would still be a valid measure of cost, the use of dollar units would be preferable. Converting the data to constant year dollars retained dollars as the unit of cost measurement and had the added benefit of retaining the variation caused by the multiplication of a random variable by a constant (reference eq.(1) page 15).

Candidate Distributions

Distribution Basis. Once data conversion was complete, a cost equation was developed to express the random variables and constants which would make up the total O & S cost estimate. That equation is

$$\text{Cost/Flying Hour} = \text{Fuel} + \text{Maintenance} + \text{Depot} \quad (2)$$

where

$$\text{Fuel} = (\text{Gallons/Flying Hour})(\text{Dollars/Gallon})$$

$$\text{Maintenance} = (\text{Manhours/Flying Hour})(\text{Dollars/Manhour})$$

$$\text{Depot} = (\text{Depot Hours/Flying Hour})(\text{Dollars/Depot Hour})$$

Each of the dollar factors are constants and the gallons, manhours, and depot hours per flying hour are random variables. In the experiment, the multiplicative aggregates, Fuel, Maintenance, and Depot were treated as random variables so as to include the increased variance caused by the constants (reference eq.(1). This was done because that variance

contributes to the true risk involved in the O & S cost estimate.

Each of the three random variables above was assumed to have a probability distribution associated with it which was represented by the VAMOS II data. To fit specific distributions to the sets of data, each set was compared in shape, central tendencies, and variance to a set of candidate distributions. The distributions used in the Monte Carlo simulation for this experiment were selected from this set of candidate distributions.

Distribution Criteria. The success of a Monte Carlo simulation for O & S costs is incumbent upon selecting cost element distributions which accurately reflect real-world costs (13:261). In an effort to achieve that accuracy, Long compiled several criteria for O & S cost element distributions (29:92). A number of the criteria pertain to the shape of the proposed distribution. It should be unimodal so that an analyst can specify a most likely value (necessary for simulation). The distribution should also be able to take on a wide variety of shapes. This characteristic allows a distribution to be 'fitted' to a set of data points like the one used in this experiment. If the distribution is flexible in shape it can be fit to both performance characteristics of systems (usually skewed left) as well as cost characteristics (usually skewed right). In addition, the parameters of the candidate distribution should be computationally simple. This would aid the computerization of the simulation.

One last criterion is not quite as clearcut as the others. The candidate distribution should have finite limits. This criterion raises important questions about the determination of a cost element

distribution. The lower (left) limit of any useful cost distribution must be finite and positive (or zero). Clearly, a cost cannot be negative in the context of O & S cost estimates. This rules out use of the Normal Distribution. It should be noted, however, that the Normal has been used extensively when other distributions have failed to match the data set. Whether the upper (right) limit should be finite or infinite is not so easily answered. A case could be made that an infinite upper limit is unrealistic. At the very least, affordability would impose a limit which is finite. There are analysts, however, that argue that the infinite upper limit accurately reflects the uncertainty present in world economics today. If the criterion of a finite upper bound is enforced, candidate distributions would be extremely limited. For the purposes of this study, a finite lower bound was deemed a necessity, but the upper bound was allowed to be infinite. A complete description of the candidate distributions follows.

Beta Distribution. The Beta distribution fits all the criteria laid out in the previous section. It can take on a wide variety of shapes, and both the lower and upper limits are finite. In addition, the distribution can be located between any two finite bounds. This distribution is particularly useful when no theoretical justification for another distribution exists (22:83). Precedence for use of the Beta for cost distributions has been set by its application in PERT networking and in Dienemann's work on cost estimation (14:15).

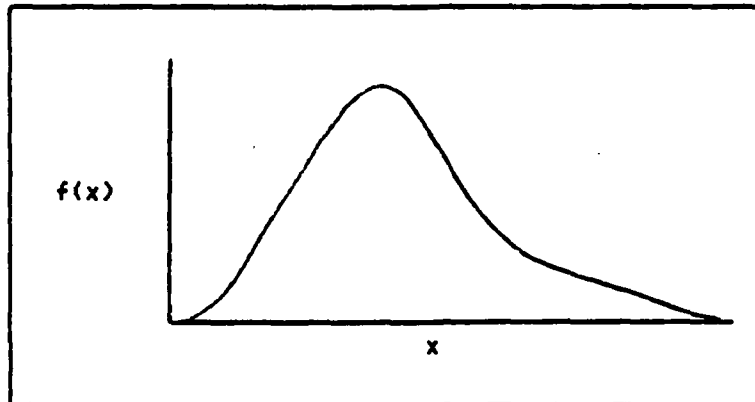


Figure 7. Example of the Beta Distribution

The Beta probability distribution function is

$$f(x) = \{ \Gamma(p+q) / \Gamma(p) \Gamma(q) \} (x-a)/b)^{p-1} (1-(x-a)/b)^{q-1}$$

where

$$\Gamma(p) = \int_0^{\infty} y^{p-1} \exp[-y] dy$$

and

p = shape parameter ($p > 0$)

q = shape parameter ($q > 0$)

a = the low value ($a > 0$)

b = the range

If the low, high, and mode for the distribution can be estimated or are known from a set of data, the parameters can be estimated using a technique developed by Donaldson (15). Coon (7) later modified this technique. This method determines a measure of the asymmetry of the distribution and assumes that the distribution is tangential to the horizontal axis at the upper and lower bounds. The decision rules developed by Donaldson and Coon using the low (L), high (H), and modal (M) values are

$$\text{If } \langle H-M \rangle > \langle M-L \rangle, \quad p = 2 \text{ and } q = \langle H-M \rangle / \langle M-L \rangle + 1 \quad (3)$$

$$\text{If } \langle M-L \rangle > \langle H-M \rangle, \quad q = 2 \text{ and } p = \langle M-L \rangle / \langle H-M \rangle + 1 \quad (4)$$

otherwise

$$p = q = 2$$

Once the parameters are known, the mean and variance can be computed:

$$\mu = p / (p + q) \quad (5)$$

$$\sigma^2 = pq / (p + q)^2 (p + q + 1) \quad (6)$$

Rectangular Distribution. The rectangular distribution can only take on a single shape, but it is computationally simple and has finite limits. The probability density function of the rectangular distribution is

$$f(x) = \begin{cases} 1 / 2h & \text{in the interval } (a-h, a+h) \\ 0 & \text{elsewhere} \end{cases}$$

where

a = the mean

The variance is

$$\sigma^2 = h^2 / 3$$

The rectangular distribution is actually a special case of the Beta distribution where $p = q = 1$. Figure 8 is a depiction of the rectangular distribution.

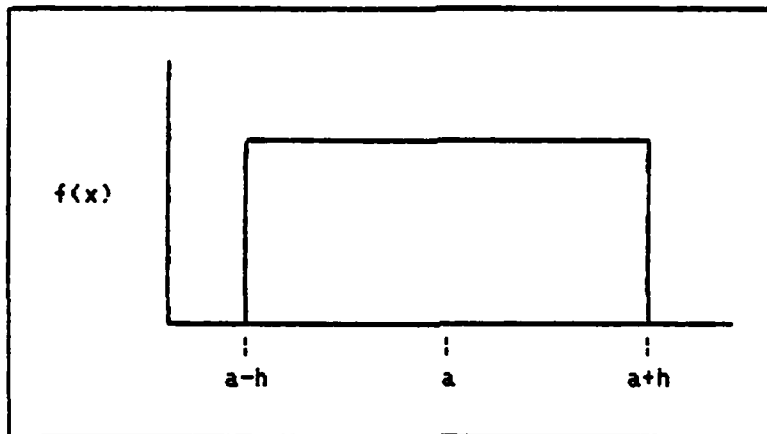


Figure 8. The Rectangular Distribution

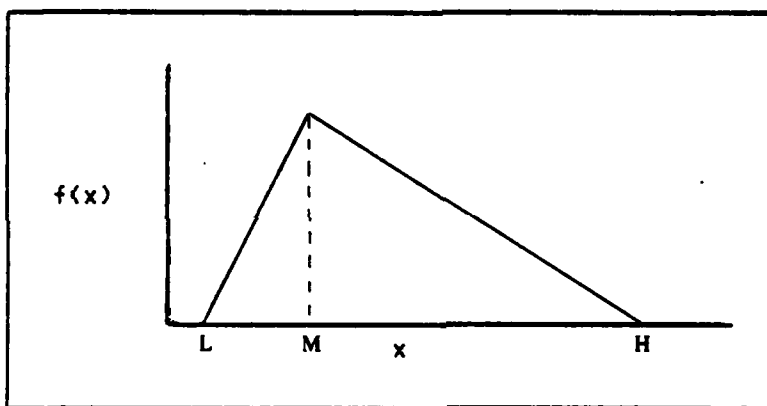


Figure 9. The Triangular Distribution

Triangular Distribution. The triangular distribution meets all the criteria for a good cost element distribution. It can take on a variety of shapes (skewness), is computationally simple, and has finite upper and lower bounds. The probability density function is

$$f(x) = \begin{cases} 2(x-L) / ((H-L)(M-L)) & L \leq x \leq M & (7) \\ 2(H-x) / ((H-L)(H-M)) & M \leq x \leq H & (8) \end{cases}$$

This distribution is completely parameterized with knowledge of L, M, and H.

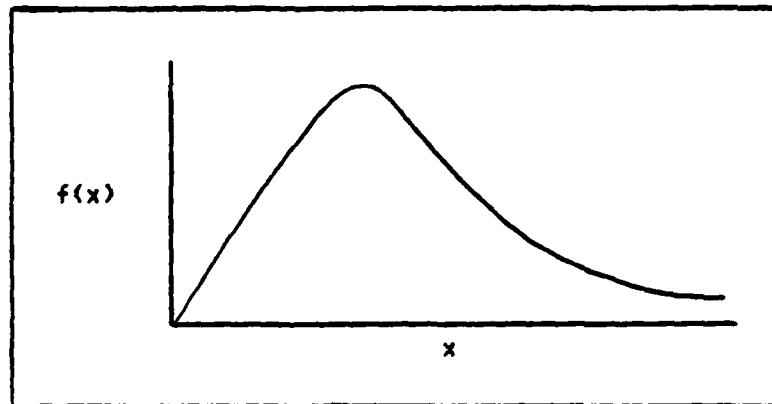


Figure 10. Example of the Gamma Distribution

Gamma Distribution. The Gamma distribution can take on a large number of shapes, but it is restricted to being skewed right because of its infinite upper bound. Like the Beta, it is useful when no theoretical justification exists for using another distribution (22:83). In addition, the infinite upper limit may reflect the true risk associated with weapon system acquisition (29:104). Figure 10 is a depiction of the Gamma distribution.

The probability density function of the Gamma distribution is

$$f(x) = (\Gamma(a))^{-1} (b)^{-a} (x)^{a-1} \exp[-x/b]$$

where

a = shape parameter

b = scale parameter

A location parameter, k, can be incorporated in the function to move the distribution along the horizontal axis. Parameterization of this distribution is somewhat involved but requires no advanced mathematics if Perry and Grieg's procedure is used (35). Their procedure requires

that an analyst's low (L) and high (H) estimates be revised to L_δ and H_δ respectively. If δ is set at .05, L now represents the value for which there is a .05 probability of anteceding it, and H the value for which there is .05 probability of exceeding it (29:105). The mean and the variance are approximated as follows:

$$\mu = (L_\delta + .95H + H_\delta) / 2.95 \quad (9)$$

$$\sigma^2 = ((H_\delta - L_\delta)^2 / 3.25) \quad (10)$$

It can also be shown that

$$\mu = ab + k \quad (11)$$

$$\sigma^2 = ab \quad (12)$$

$$\text{mode} = M = b(a - 1) + k \quad (13)$$

By subtraction (eq (11) - eq (13))

$$b = \mu - M$$

Equation (12) can be used to solve for a, and equation (13) can be used to solve for k.

Selection and Validation. Although the Log-Normal and Weibull could also have been candidates, they were eliminated from consideration because of some associated difficulties. The Log-Normal is conditional on $\log(x-a)$ having a Normal distribution. This did not appear to be true for the data obtained from VAMOS II. The Weibull was eliminated because the Gamma distribution appears to be able to take on all the same shapes (those relevant to O & S cost estimation) and parameterization was easier. The Normal distribution, as previously stated, was eliminated because of its infinite lower bound.

Relative frequency histograms of the extracted data provided an initial shape from which to judge the relevance of the candidate

distributions. At this point the rectangular distribution was dropped as a candidate simply because all three data sets (fuel, maintenance, and depot) showed definite peaks (modes) in their histograms. Appendix B contains the parameter calculations, and Appendices E to G contain the data set histograms.

Parameterization was then accomplished using the low, high and modal values for each of the three data sets. For both fuel and depot cost elements all three candidate distribution parameters could be determined; however, the Gamma distribution parameters could not be calculated for the maintenance element. The maintenance data showed a definite skew to the left which the Gamma distribution cannot accomodate.

To further validate the use of the 'fitted' distributions, a Kolmogorov two-sided test was conducted. This test measures the maximum difference between a set of data points and a theoretical distribution (6:301). Each cost element data set was compared to each of the distributions fitted to it via parameterization. The hypotheses were

$$H_0 : S(x) = F(x)$$

$$H_1 : S(x) \neq F(x)$$

where

$S(x)$ = the area under the curve of the baseline data

$F(x)$ = the area under the curve of a theoretical distribution

x = each specific data point

When $F(x)$ was the Triangular distribution, its values were determined using the calculations of Appendix D. When $F(x)$ was either the Gamma or Beta distribution, values were extracted from tables (23:6-8,166-172).

For the Gamma distribution, values for $F(x)$ could be extracted directly, although interpolation was required at times. For the Beta distribution, values had to be extracted from a percentage points table and graphed because the tables did not contain all the points of the Beta distribution. The value of $F(x)$ was then extracted from that graph. The test required computation of the difference between $F(x)$ and $S(x)$ for each data point, and comparison of the maximum difference to a test statistic developed by Kolmogorov. Appendix H contains a computer program, written in Basic, to perform the calculations for the test. Chapter III contains a detailed discussion of the results of this test.

Monte Carlo Simulation

Justification. Based on the CAIG CES, O & S cost for a weapon system contains eight major categories of cost and a multitude of subcategories. Each is a random variable, and when a large number of random variables is involved, simulation is an ideal medium for modelling the system (22:329). Monte Carlo simulation is versatile, requires no complex math, and allows the analyst to alter the system to determine the effects of various situations. On the other hand, simulation is prone to sampling errors, is limited to situations chosen by the analyst, and is only valid for the specific situation modelled (13:262-263). Simulation requires expression of a random variable as a probability distribution, and results in a quantification of the degree of uncertainty (14:5).

There are many precedents for using Monte Carlo simulation in O & S cost estimating. Both Dienemann (14) and Worm (39) have done extensive

experimentation in this area. Further, since quantitative risk analysis requires use of stochastic methods, simulation is an ideal methodology. The Monte Carlo simulation used in this experiment provided a means to view, then judge the effects of various assumptions about O & S cost elements.

Program Structure. The simulation was written in Fortran 77. A listing of the program is included in Appendix C. Using the data extracted from VAMOS II, the cost equations developed above, and the parameterization techniques previously detailed, a simulation was devised which would vary the distributions for each cost element over the selected distributions, sample those distributions, and sum the samples. Cost estimating by Monte Carlo simulation is not normally performed in precisely this manner. Usually, the program is interactive, soliciting a low, high, and most likely estimate from the analyst. The distributions to be used are fixed, and the formulae for parameterization are included in the source code. In this way, an analyst can easily see the effect his estimates of the low, high, and mode have on the total cost. Since the focus of this research is, however, the effect of assumptions about the cost element distributions used, this experiment used a database to determine the low, high, and mode, fixed them, and varied the distributions. In this manner eighteen different total cost distributions were derived. Table I is a summary of the sampling combinations.

The program made use of subroutines, developed by IMSL Inc., to sample the distributions.

TABLE I

Sampling Combinations

Combination	Distribution Sampled		
	Fuel	Maint.	Depot
1	Triangular	Triangular	Triangular
2	"	"	Gamma
3	"	"	Beta
4	"	Beta	Triangular
5	"	"	Gamma
6	"	"	Beta
7	Gamma	Triangular	Triangular
8	"	"	Gamma
9	"	"	Beta
10	"	Beta	Triangular
11	"	"	Gamma
12	"	"	Beta
13	Beta	Triangular	Triangular
14	"	"	Gamma
15	"	"	Beta
16	"	Beta	Triangular
17	"	"	Gamma
18	"	"	Beta

For the Beta and Gamma distributions, the IMSL routines provided a complete sample space; however, these routines are constructed to sample from the most general form of each distribution. For example, the Beta subroutine GGBTR, samples from a Beta distribution located between zero and 1 on the horizontal axis. Consequently, it was necessary to convert the locations of the samples. The source code for the conversions follows the call for the IMSL subroutines (reference Appendix C).

The procedure for sampling from the Triangular distribution was not quite so straightforward. An IMSL subroutine (GGUBS) was used to sample a uniform distribution. Those values represented the area under the curve for a uniform deviate. In order to derive a triangular deviate representing the same area the following integrals had to be solved for x .

$$F(x) = \int_L^x (2(x-L)) / ((H-L)(M-L)) dx \quad 0 \leq F(x) \leq U$$

and

$$F(x) = \int_x^H (2(H-x)) / ((H-L)(H-M)) dx + U \quad U < F(x) \leq 1$$

where

$F(x)$ = the uniform deviate

U = the area under the curve from L to M

The IMSL routines make use of a random seed, 'DSEED', to generate distribution deviates. If the same seed is used to sample two distributions, the random numbers generated will be identical. There are three basic ways to treat this variable in O & S cost simulations. For a given sampling combination, the same seed can be used across all cost elements. This method was used by Dienemann (14) and is generally reserved for situations where the cost elements are considered to be

completely dependent. For this simulation, complete independence was assumed. The disadvantage of using Dienemann's technique is that it implies that if one element cost is high the other element costs will be high as well. This has a tendency to magnify the risk (14:4-5). The second treatment of the seed is to allow it to differ for each sample taken. This is based on element independence, and it is the most common treatment. The third treatment, used in this experiment, is to hold the seed constant for each cost element regardless of the distribution sampled. This treatment was developed to eliminate differences in the resultant total cost vectors due to random number generation. In this way, the changes in the resultant totals could be traced solely to the different distributions sampled.

The method of building the simulation followed the accepted procedures (13:253). The problem was put in quantitative form, the model was constructed, empirical data was obtained for validation, the source code was written and tested, the model was validated, and the results were evaluated/analyzed. Analysis of the validation and the simulation is contained in Chapter III.

III. Results and Findings

The results of this experiment can be broken into two categories: 1) the result of validation of the simulation, and 2) the results of the simulation. Each category will be discussed and then the simulation results will be applied to the research questions posed in Chapter I.

Validation Results

Validation of a Monte Carlo simulation model is critical. It insures that the model, in fact, represents the situation being investigated (13:254). In this experiment, the validation centered on whether or not the samples drawn reflected the baseline data sets for each of the cost elements. The distributions were compared visually, by computer generated measures, and by the Kolmogorov two-sided test. The sampling routines in the simulation were not validated because IMSL had done so previously (28).

Visual comparison of the baseline cost element histograms and the sample histograms confirmed an apparent reflection of the baseline data sets. The histograms for each baseline data set and the samples corresponding to that data set are contained in Appendices E through G for the fuel, maintenance, and depot cost elements respectively. Although the histograms of the samples showed some evidence of skew not present in the baseline histograms, this was to be expected. Minor differences for all the samples can be attributed to sampling error. Even when 1000 iterations are used, samples will never perfectly reflect

the distribution from which they have been drawn. Additionally, the samples from the Gamma distribution for both the fuel and depot elements were skewed further right than the baseline sets. This was to be expected because a distribution with an infinite upper bound (the Gamma) was parameterized by a data set with a finite upper bound. The parameterization of the Gamma involved redefining the finite upper bound of the baselines as the value which represents the 95th percentile of the Gamma distribution (reference page 37). It should be noted here that the histograms used were generated by the 'S' statistics package on the Univac 11/780 computer. The intervals for all the histograms used in this experiment were machine generated unless there was a specific need to delineate them in a particular manner.

The means, variances, modes and ranges of the baseline data sets and the samples were generated by the computer. Tables II and III are summaries of that information. The differences between baseline and sample measures in these two tables was attributed to the nature of sampling and to approximations used in parameterization. There were no unexpected differences in the data. The samples cannot be expected to precisely echo the baselines because they were developed using a Monte Carlo simulation and thus, not all points in a continuous distribution will be present. Likewise, the parameterizations are approximations. They do not precisely fit a data set to a distribution. One distinct problem with the parameterization method used was the determination of the mode.

TABLE II

Comparison of Means and Modes

Cost Element	Distribution	Mean		Mode	
		Data	Sample	Data	Sample
Fuel	Triangular	1862.52	1874.71	1700.00	1818.85
	Gamma		1858.88		1717.63
	Beta		1781.48		1631.80
Maintenance	Triangular	1315.70	1324.26	1400.00	1344.77
	Beta		1338.05		1498.42
Depot	Triangular	1302.78	1393.01	1100.00	1428.95
	Gamma		1446.28		1045.74
	Beta		1248.32		1069.39

TABLE III

Comparison of Variances and Ranges

Cost Element	Distribution	Variance		Range	
		Data	Sample	Data	Sample
Fuel	Triangular	60076.4	47601.2	1067.41	1001.93
	Gamma		103192.4		1844.87
	Beta		38120.2		928.72
Maintenance	Triangular	96827.7	31563.7	1375.42	1146.20
	Beta		83418.1		1333.22
Depot	Triangular	254684.7	56673.8	1708.19	1460.16
	Gamma		298634.0		3574.33
	Beta		85006.8		1486.83

The centerpoint of the modal interval on the baseline histograms was used as the distribution mode. However, if the intervals were altered, the mode would shift higher or lower. This would not be a problem if analyst estimates of the low, high, and mode were used to calculate the parameters; therefore, the fact that the mode could shift was ignored.

The Kolmogorov test was performed to determine whether or not the assumed distribution reflected the data. The test compared the area under the curve of the data set for a given cost 'x' ($S(x)$) to the area under the curve of a theoretical distribution for the same point ($F(x)$). In this test the theoretical distributions were the Triangular, the Beta, and the Gamma. The maximum difference was compared to the test statistic developed by Kolmogorov (E1:296). Based on this comparison, the null hypothesis (H_0) was either rejected at a specific confidence level (α) or not rejected. The results of the test are compiled in Table IV. Four of the eight null hypotheses were rejected at a relatively high confidence level, including two of the Beta distributions. The high maximum differences for the Beta distributions may have been due to the method used to extract the values for $F(x)$. The graphs may not have been precise enough for the test because of the large amount of interpolation required to develop the graphs and to extract the values from them. However, the maximum differences were large enough to indicate that interpolation error did not affect the net outcome. Rejection of the Triangular and Gamma distributions for the Depot cost element was not surprising.

TABLE IV

Kolmogorov Test Results

Cost Element	Null Hypothesis	Maximum Difference	Implication
Fuel	$S(x) = \text{Triangular}$.0377	Do not reject H_0
Fuel	$S(x) = \Gamma(x; 2, 3.68)$.3074	Reject H at $\alpha = .01$
Fuel	$S(x) = \beta(x; 3.96)$.0990	Do not reject H_0
Maintenance	$S(x) = \text{Triangular}$.0770	Do not reject H_0
Maintenance	$S(x) = \beta(x; 2.42, 2)$.1066	Do not reject H_0
Depot	$S(x) = \text{Triangular}$.2961	Reject H at $\alpha = .01$
Depot	$S(x) = \Gamma(x; 2, 4.5)$.2580	Reject H at $\alpha = .01$
Depot	$S(x) = \beta(x; 2.67)$.1886	Reject H at $\alpha = .1$

The high degree of skewness evident in the histogram of the Depot data set indicated that it would be difficult to fit a distribution to it. In both cases, the maximum difference occurred in the region just above the mode, where the data curve dropped off rapidly (see Appendix G). The theoretical Triangular and Gamma distributions do not exhibit this characteristic.

The fact that fully half of the samples did not reflect the baseline data sets would be critical if that was the intent of the experiment. However, for this effort, the data sets were used only as a surrogate for analyst estimates of the low, mode, and high values for a cost element. Typically, analysts do not use an historical data set for

those estimates. Further, the small amount of data available implies that the data set may not truly reflect the true distribution of the cost element. It is important to note, though, that as the O & S cost database increases, it will come much closer to reflecting actual cost element distributions. As that reflection becomes more precise, far more effort and rigor will be required to model the cost distributions.

The results of the simulation validation did not refute the model. Certainly, this was not absolute proof that the model was precise, but in the absence of further negative information, the model was considered to be valid.

Simulation Results

The final product of the simulation was an 1000 row, 18 column array of total costs, each column being the aggregation of three cost elements. Each column contained 1000 data points for a yearly cost per flying hour distribution. The differences in variance of these distributions reflected the effect of assumptions about cost element distributions on cost estimate risk analysis.

The variance of each of the eighteen resultant distributions was computed with the aid of the 'S' statistics package, on the Univac 11/780 computer. Table V is a rank order of the distributions' variance (highest to lowest).

TABLE V

Rank-Ordered Variances

Rank	Cost Element Distributions			Variance
	Fuel	Maintenance	Depot	
1	Gamma	Beta	Gamma	484456.5
2	Gamma	Triangular	Gamma	449804.0
3	Triangular	Beta	Gamma	422026.0
4	Beta	Beta	Gamma	417061.6
5	Triangular	Triangular	Gamma	374132.0
6	Beta	Triangular	Gamma	363732.0
7	Gamma	Beta	Beta	256668.2
8	Gamma	Beta	Triangular	226279.0
9	Gamma	Triangular	Beta	224676.5
10	Triangular	Beta	Beta	210462.2
11	Beta	Beta	Beta	208728.2
12	Triangular	Beta	Triangular	195089.7
13	Gamma	Triangular	Triangular	190589.6
14	Beta	Beta	Triangular	185032.4
15	Triangular	Triangular	Beta	165229.4
16	Beta	Triangular	Beta	158059.4
17	Triangular	Triangular	Triangular	146159.2
18	Beta	Triangular	Triangular	130666.0

It is apparent that the presence of a Gamma distribution in a combination had a definite effect on the total variance of that combination. There were ten combinations which included the Gamma distribution, and nine of these were the nine highest variance combinations. Below the ninth rank, however, there was no discernable pattern to the order.

There was also a relationship established between the rank order and the cost element which was Gamma distributed. There were six combinations in which the Depot cost element was a Gamma distribution. These six combinations were the top six in the rank order. Table VI highlights this point.

TABLE VI

Gamma Distribution Effects

Rank	Gamma Distributed Cost Element(s)
1	Depot, Fuel
2	Depot, Fuel
3	Depot
4	Depot
5	Depot
6	Depot
7	Fuel
8	Fuel
9	Fuel
13	Fuel

When the Fuel cost element was Gamma distributed, the cost estimate variance was also consistently high.

Findings

Research Question 1. The first question posed was: Do the assumptions about cost element distributions affect the risk of an O & S cost estimate? The evidence suggests that the assumption about cost element distributions is critical to risk analysis. If the analyst assumes a distribution for a given cost element in lieu of another, the risk of the total cost estimate will have a different magnitude. Further, if distributions with infinite upper bounds are assumed, it seems likely that the magnitude of variance will be higher than if finite distributions are assumed. This experiment provides proof of this but only for the cost elements used and only for the Gamma distribution as opposed to the Beta or Triangular. By extrapolation,

this same effect should hold true for other distributions with infinite upper bounds such as the Weibul or Log-Normal.

The implication of this conclusion is that risk analysts may unwittingly impact the risk of a cost estimate through assumptions about cost element distributions. For example, an estimate based on all Beta distributions (such as Dienemann (14) used) would provide a conservative (lower) risk estimate, whereas, an estimate based on all Gamma distributions would have a relatively large risk associated with it. It is not clear from this experiment whether or not there is a similar effect when two estimates are both based on finite, but different, distributions. It does, however, clearly indicate that each cost element distribution combination yields a unique cost estimate variance and subsequent risk.

Research Question 2. The second question asked whether or not the effect of assumptions was consistent for all cost elements. The evidence collected in this effort indicated that there was no consistency among the cost elements for specific distributions. On the other hand, there appeared to be a strong suggestion that finite and infinite distributions do have a consistent effect.

To establish consistency (or lack thereof) the variance of the cost estimate was measured while holding two element distributions constant and allowing the third to vary. This was done using the data compiled in Table III. For example, if the Depot and Maintenance element distributions are held constant as Gamma and Beta respectively, the Gamma distribution for the Fuel element yields the highest variance with the Triangular next, and the Beta the lowest. This order was constant

regardless of the Depot and Maintenance combination. The Gamma, then Beta, then Triangular was the order for the Depot element, while the order for the Maintenance element was Beta then Triangular. The effect of the Gamma distribution appeared consistent, but there was no clear distinction between the Triangular and Beta distributions.

Research Question 3. Question three asked: Can actual cost element distributions be determined with data presently available? The experiment was not designed to provide empirical results with which to answer this question. Instead, the background research and investigation provided the answer.

The VAMOS II database is in its infancy. As such, it can only provide a limited number of datapoints, and the accuracy of those points is, as yet, undetermined. Even if VAMOS II could provide several datapoints, there would be no method to absolutely determine the cost element distributions. Theoretically, it would require an infinite number of datapoints; however, as the database grows, the possible distributions for each cost element should decline in number. This will limit analyst assumptions and may provide estimates with a more accurate degree of risk.

Follow-On Research

As with any research effort, this project uncovered several topics which need to be investigated and reported. The topics fall into three basic areas: 1) problems associated with databases, 2) problems in other areas of risk analysis, and 3) continuation of this effort.

A real-world, accurate cost database is essential to further Air

Force efforts in improving risk analysis. VAMOS II is certainly a major step in that direction, but there are still a multitude of unanswered questions about that database. It is subject to data collection errors as are all databases. Efforts to determine the magnitude and distribution of erroneous data input to the VAMOS II algorithms would significantly enhance that database. Data trends also need to be studied. For example, during this research it was noted that the Fuel cost element data had a noticeable trend. The newer fighter aircraft have lower fuel costs (usage) than older aircraft of the same type. This trend is consistent over all the fighter aircraft in the active inventory. This may reflect the new engine technology of recent years. Time series analysis of the data may confirm or refute this observation. If fuel cost (usage) is actually moving toward lower values, cost analyst estimates of the most likely value for the cost element (the mode) would be affected. That estimate is critical in Monte Carlo simulation of O & S costs.

Risk analysis methodology is also an area where research would be very valuable. Further investigation of the transform method of risk analysis as presented by Long (29) may result in a precise determination of cost estimate distribution and risk. Such a method would eliminate the sampling errors and constraints associated with Monte Carlo simulation. The key is to determine a method to easily and accurately invert and reinvert transforms. If that could be done and be computerized, a much more precise method of cost estimate risk analysis would be available. On a more qualitative note, investigation of the methods of applying and presenting risk analysis in the Air Force

acquisition process would be valuable. In the interim, however, simulation appears to be the most feasible risk analysis methodology, and there are a number of ways to improve it.

The experiment in this study was simplex and did not involve a large number of variables. Using this methodology, research could be conducted using samples of several thousand or more cost element distributions. Broadening this area of study may provide results which can be used to further delineate the ramifications of assumptions about cost element distributions. Also, a study to determine the adequacy of parameterization methods would provide useful information.

Summary

Risk analysis is not a panacea for the problems associated with cost estimation. It is but one tool for the decision-maker to use. There is absolutely no way to precisely measure the risk of a decision, only ways to quantify portions of the risk.

...as long as uncertainty is present, it is possible to have outcomes with unfavorable consequences even when 'good' decisions have been made [24:7].

There is no resolution to this dilemma. It is incumbent upon Air Force decision-makers to accept that fact yet continue to work at improving cost estimation.

Appendix A: Data and Constant Year Dollars Factors

FY 1981 Cost Data (Constant 1983 Dollars per Flying Hour)

<u>Aircraft</u>	<u>Flying Hours</u>	<u>Fuel \$/hr</u>	<u>Maint \$/hr</u>	<u>Depot \$/hr</u>
FB-111A	17232.00	1409.73	1230.14	1433.37
F-4C	8947.00	1690.26	1967.65	1274.14
F-4D	69246.00	1769.51	1409.93	969.60
F-4E	109778.00	1826.42	1422.18	1115.80
F-4G	24335.00	1770.12	1074.85	978.09
RF-4C	46091.00	1468.02	1029.24	720.57
F-15A	70456.00	1623.11	1275.87	947.44
F-15B	12418.00	1445.60	938.45	828.16
F-15C	35394.00	1641.49	1104.59	800.22
F-15D	5931.00	1570.92	997.95	848.24
F-111A	17830.00	1765.15	1559.43	1280.52
F-111D	17552.00	1639.25	1461.78	1630.89
F-111E	19095.00	1843.67	592.23	1561.67
F-111F	21977.00	1878.49	1109.89	1960.86

Conversion Factors

Fuel Cost Element: FY 81 Data / 1.132 = FY 83 Dollars

Maintenance Cost Element: FY 81 Data / .875 = FY 83 Dollars

Depot Cost Element: FY 81 Data / .873 = FY 83 Dollars

FY 1982 Cost Data (Constant 1983 Dollars per Flying Hour)

<u>Aircraft</u>	<u>Flying Hours</u>	<u>Fuel \$/hr</u>	<u>Maint \$/hr</u>	<u>Depot \$/hr</u>
FB-111A	15904.00	1605.01	1419.13	1653.74
F-4C	7140.00	1939.91	1891.19	1093.70
F-4D	45405.00	1951.83	1488.25	903.22
F-4E	104140.00	2102.27	1690.30	1081.13
F-4G	24160.00	2134.97	1215.81	919.02
RF-4C	44379.00	1692.70	1380.18	744.61
F-15A	72323.00	1785.75	1306.78	922.77
F-15B	13336.00	1726.74	823.51	782.42
F-15C	48948.00	1997.34	1078.74	775.95
F-15D	9503.00	1833.87	864.31	762.23
F-111A	16308.00	2050.30	1736.43	1542.64
F-111D	16667.00	1788.33	1763.34	1708.11
F-111E	18649.00	2282.26	1305.87	2271.99
F-111F	22158.00	2287.03	1414.79	2428.76

Conversion Factors

Fuel Cost Element: FY 82 Data / 1.114 = FY 83 Dollars

Maintenance Cost Element: FY 82 Data / .96 = FY 83 Dollars

Depot Cost Element: FY 82 Data / .953 = FY 83 Dollars

FY 1983 Cost Data

<u>Aircraft</u>	<u>Flying Hours</u>	<u>Fuel \$/hr</u>	<u>Maint \$/hr</u>	<u>Depot \$/hr</u>
FB-111A	17736.00	1622.12	1297.42	2014.60
F-4C	---	---	---	---
F-4D	31768.00	1965.06	1537.33	1180.84
F-4E	115797.00	2151.61	1545.81	1191.00
F-4G	23755.00	2265.97	1189.90	1269.75
RF-4C	46838.00	1713.99	1249.73	1108.91
F-15A	71908.00	1848.31	1501.53	1040.19
F-15B	14130.00	1701.20	1034.89	1023.78
F-15C	62031.00	1902.82	1087.28	1024.26
F-15D	12445.00	2477.14	843.87	1005.06
F-111A	14410.00	2104.23	1802.78	1973.28
F-111D	17927.00	2068.78	1733.53	2294.70
F-111E	20110.00	2027.40	1237.74	1980.41
F-111F	22059.00	1994.79	1328.98	2367.38

Note: For all years, raw data can be obtained by multiplying the data above by the number of flying hours for the aircraft.

Appendix B: Cost Element Parameterization

Fuel Cost Element

$$\begin{aligned}L &= 1409.73 \\M &= 1700.00 \\H &= 2477.14\end{aligned}$$

Beta Distribution Parameters

Using the decision rules on page 33:

$$H - M = 2477.14 - 1700.00 = 777.14$$

$$M - L = 1700.00 - 1409.73 = 290.27$$

therefore

$$p = 2$$

$$q = (H - M) / (M - L) + 1 = (777.14 / 290.27) + 1 = 3.68$$

Gamma Distribution Parameters

Using the formulae on page 37:

$$\begin{aligned}\mu &= (L_0 + .95M + H_0) / 2.95 \\&= (1409.73 + .95(1700) + 2477.14) / 2.95 \\&= 1865.04\end{aligned}$$

$$\begin{aligned}\sigma^2 &= ((H - L) / 3.25)^2 \\&= (1067.14 / 3.25)^2 \\&= 107868.79\end{aligned}$$

$$B = \mu - M = 1865.04 - 1700 = 165.04$$

$$A = \sigma^2 / B = 107868.79 / 165.04 = 3.96$$

$$K = \mu - AB = 1865.04 - (3.96)(165.04) = 1211.48$$

Maintenance Cost Element

$$\begin{aligned}L &= 592.23 \\M &= 1400.00 \\H &= 1967.65\end{aligned}$$

Beta Distribution Parameters

Using the decision rules on page 33:

$$H - M = 1967.65 - 1400.00 = 567.65$$

$$M - L = 1400.00 - 592.23 = 807.77$$

therefore

$$q = 2$$

$$p = (M - L) / (H - M) + 1 = (807.77 / 567.65) + 1 = 2.42$$

Gamma Distribution Parameters

Using the formulae on page 37:

$$\begin{aligned}\mu &= (L_{\delta} + .95M + H_{\delta}) / 2.95 \\&= (592.23 + .95(1400) + 1967.65) / 2.95 \\&= 1318.60\end{aligned}$$

$$B = \mu - M = 1318.60 - 1400 = -81.4$$

Because B is negative, this distribution cannot be used for the Maintenance Cost Element

Depot Cost Element

L = 720.57
M = 1100.00
H = 2428.76

Beta Distribution Parameters

Using the decision rules on page 33:

$$H - M = 2428.76 - 1100.00 = 1328.76$$

$$M - L = 1100.00 - 720.57 = 379.43$$

therefore

$$p = 2$$

$$q = (H - M) / (M - L) + 1 = (1328.76 / 379.43) + 1 = 4.5$$

Gamma Distribution Parameters

Using the formulae on page 37:

$$\begin{aligned}\mu &= (L + .95M + H) / 2.95 \\ &= (720.57 + .95(1100) + 2428.76) / 2.95 \\ &= 1421.81\end{aligned}$$

$$\begin{aligned}\sigma^2 &= ((H - L) / 3.25) \\ &= (1708.19 / 3.25) \\ &= 276252.13\end{aligned}$$

$$B = \mu - M = 1421.81 - 1100 = 321.81$$

$$A = \sigma^2 / B = 276252.13 / 321.81 = 2.67$$

$$K = \mu - AB = 1421.81 - (2.67)(321.81) = 562.58$$

Appendix C: Monte Carlo Simulation Source Code

PROGRAM MONTE

VARIABLES

NR.....NUMBER OF SAMPLES DRAWN
 WK.....WORKSPACE VECTOR OF 2 * NR
 R.....VECTOR OF DEVIATES
 DSEED.....RANDOM NUMBER SEED
 LO,MODE,HI...PRECEDED BY AN F, M, OR D, INDICATE
 VALUES FOR THAT DATA SET: F=FUEL,
 M=MAINTENANCE, D=DEPOT
 P, Q.....BETA DISTRIBUTION PARAMETERS
 A.....GAMMA DISTRIBUTION PARAMETER
 CV.....AREA UNDER THE CURVE OF THE TRIANGULAR
 DISTRIBUTION BELOW THE MODE
 SAMPLE.....ARRAY OF SAMPLES FROM THE THREE
 DISTRIBUTIONS FOR EACH OF THE THREE
 COST ELEMENTS
 TCOST.....ARRAY OF SUMS OF SAMPLE COSTS

INITIALIZATION

INTEGER NR
 REAL R(1001), FLO, MLO, DLO, FMODE, MMODE, DMODE
 REAL FHI, MHI, DHI, P, Q, A, TCOST(1000,18),WK(2000)
 REAL SAMPLE(1000,8), B, C, CV
 DOUBLE PRECISION DSEED

FUEL SAMPLES

NR = 1000
 FLO = 1409.73
 FMODE = 1700.00
 FHI = 2477.14
 A = 3.96
 B = 165.04
 C = 1211.48
 P = 2.0

C

C

20 CONTINUE

30 CONTINUE

C

C

2

3

C

C

C

```

C *****
C
C      DEPOT SAMPLES
C      *****
C
C      DLO = 720.57
C      DMODE = 1100.00
C      DHI = 2428.76
C      A = 2.67
C      B = 321.81
C      C = 562.58
C      P = 2.0
C      Q = 4.50
C      CV = 0.222124
C
C      DSEED = 7326.D0
C      CALL GGTRA(DSEED,NR,R)
C      DO 70 I = 1,1000
C         IF (R(I).LE.CV) THEN
C            SAMPLE(I,6) = DLO + SQRT(R(I)*(DHI-DLO)*(DMODE-DLO))
C         ELSE
C            SAMPLE(I,6) = DHI - SQRT((1-R(I))*(DHI-DLO)*(DHI-DMODE))
C         ENDIF
C      70 CONTINUE
C
C      DSEED = 7326.D0
C      CALL GGAMR(DSEED,A,NR,WK,R)
C      DO 80 I = 1,1000
C         SAMPLE(I,7) = C + R(I)*B
C      80 CONTINUE
C
C      DSEED = 7326.D0
C      CALL GGBTR(DSEED,P,Q,NR,R)
C      DO 90 I = 1,1000
C         SAMPLE(I,8) = DLO + R(I)*(DHI - DLO)
C      90 CONTINUE
C
C *****
C
C      SUMS FOR TOTAL COST ARRAY
C      *****
C
C      H = 0
C      DO 110 J = 1,3
C         DO 110 K = 4,5
C            DO 110 L = 6,8
C               H = H + 1
C               DO 110 I = 1,1000
C                  TCost(I,H) = SAMPLE(I,J) + SAMPLE(I,K) + SAMPLE(I,L)
C            110 CONTINUE
C
C *****
C

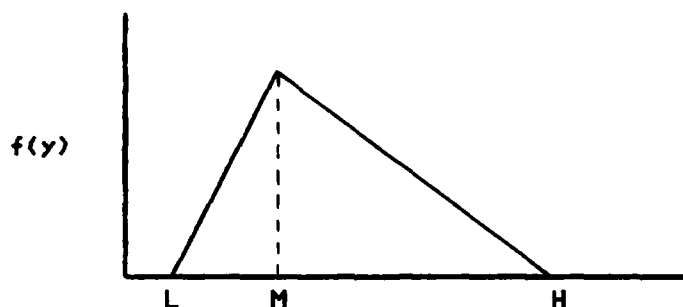
```

```

C      PRINT RESULTS TO EXTERNAL FILE
C      *****
C
C      DO 120 I = 1,1000
C          WRITE (15,200)(TCOST(I,J),J=1,18)
C          WRITE (16,210)(SAMPLE(I,J),J=1,8)
C      120 CONTINUE
C
C      200 FORMAT(' ',18(1X,F7.2))
C      210 FORMAT(' ',8(1X,F7.2))
C
C      END
C
C      *****

```

Appendix D: Triangular Distribution Calculations



$$f(y) = \begin{cases} \frac{2(y-L)}{(H-L)(M-L)} & L \leq y \leq M \quad (7) \\ \frac{2(H-y)}{(H-L)(H-M)} & M \leq y \leq H \quad (8) \end{cases}$$

where

L = low value

M = mode

H = high value

Area Below the Mode ($L \leq y \leq M$)

$$\begin{aligned} \text{Area} &= \int_L^y \frac{2(x-L)}{(H-L)(M-L)} dx \\ &= \frac{2}{(H-L)(M-L)} \int_L^y (x-L) dx \\ &= \frac{2}{(H-L)(M-L)} \left[\left(\frac{y^2 - L^2}{2} \right) - L(y-L) \right] \\ &= \frac{(y-L)^2}{(H-L)(M-L)} \end{aligned}$$

Area Above the Mode ($M < y \leq H$)

Since the total area under the curve = 1:

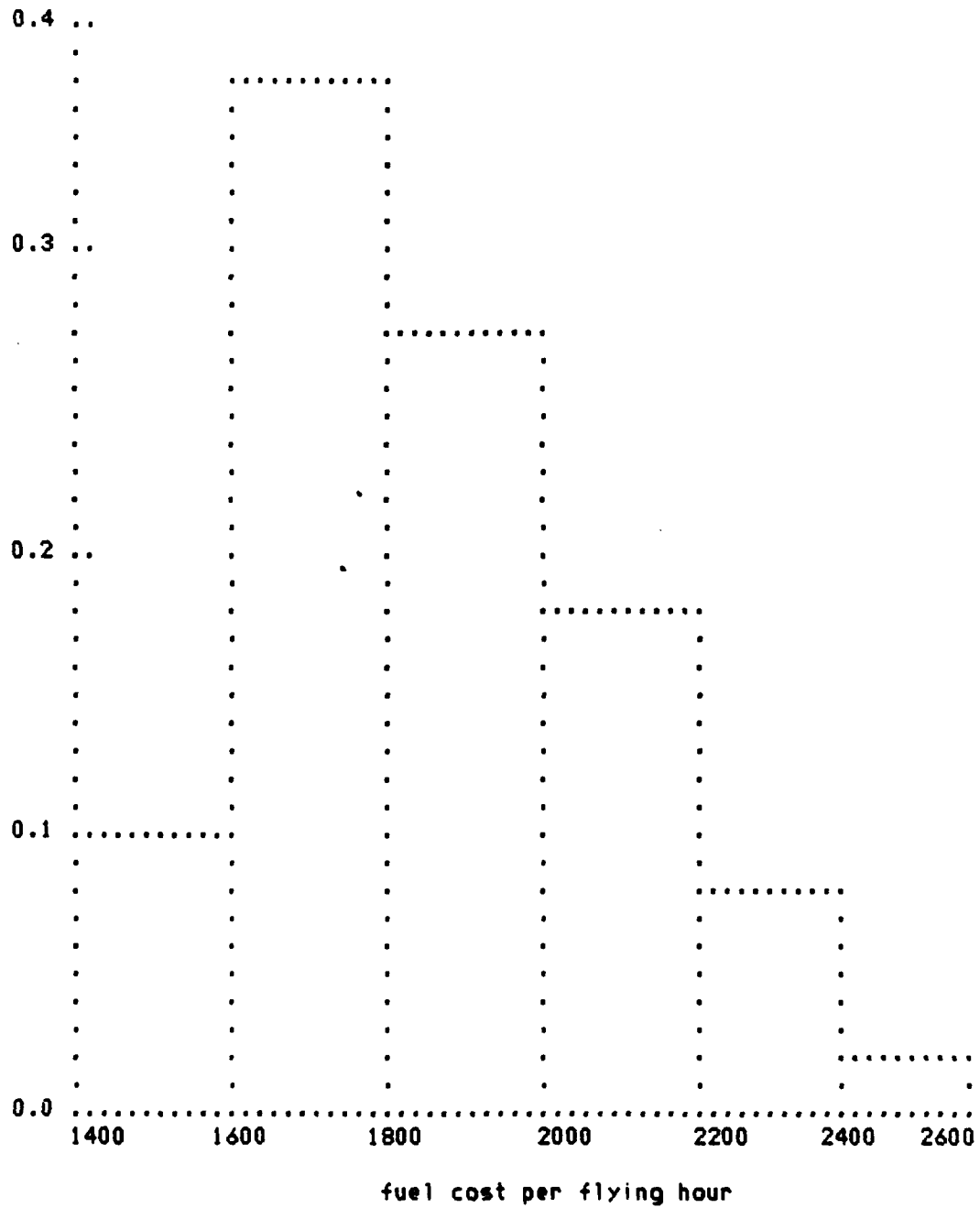
$$\begin{aligned} \text{Area} &= 1 - \int_y^H \frac{2(H-x)}{(H-L)(H-M)} dx \\ &= 1 - \frac{2}{(H-L)(H-M)} \int_y^H (H-x) dx \\ &= 1 - \left[\frac{2}{(H-L)(H-M)} \right] \left[(H^2 - Hy) - \frac{(H^2 - y^2)}{2} \right] \\ &= 1 - \left[\frac{(H-y)^2}{(H-L)(H-M)} \right] \end{aligned}$$

Substitution of the values for L, M, and H for each of the cost elements into the equations above yielded the values for the variable 'CV' used in the Monte Carlo simulation (see Appendix C).

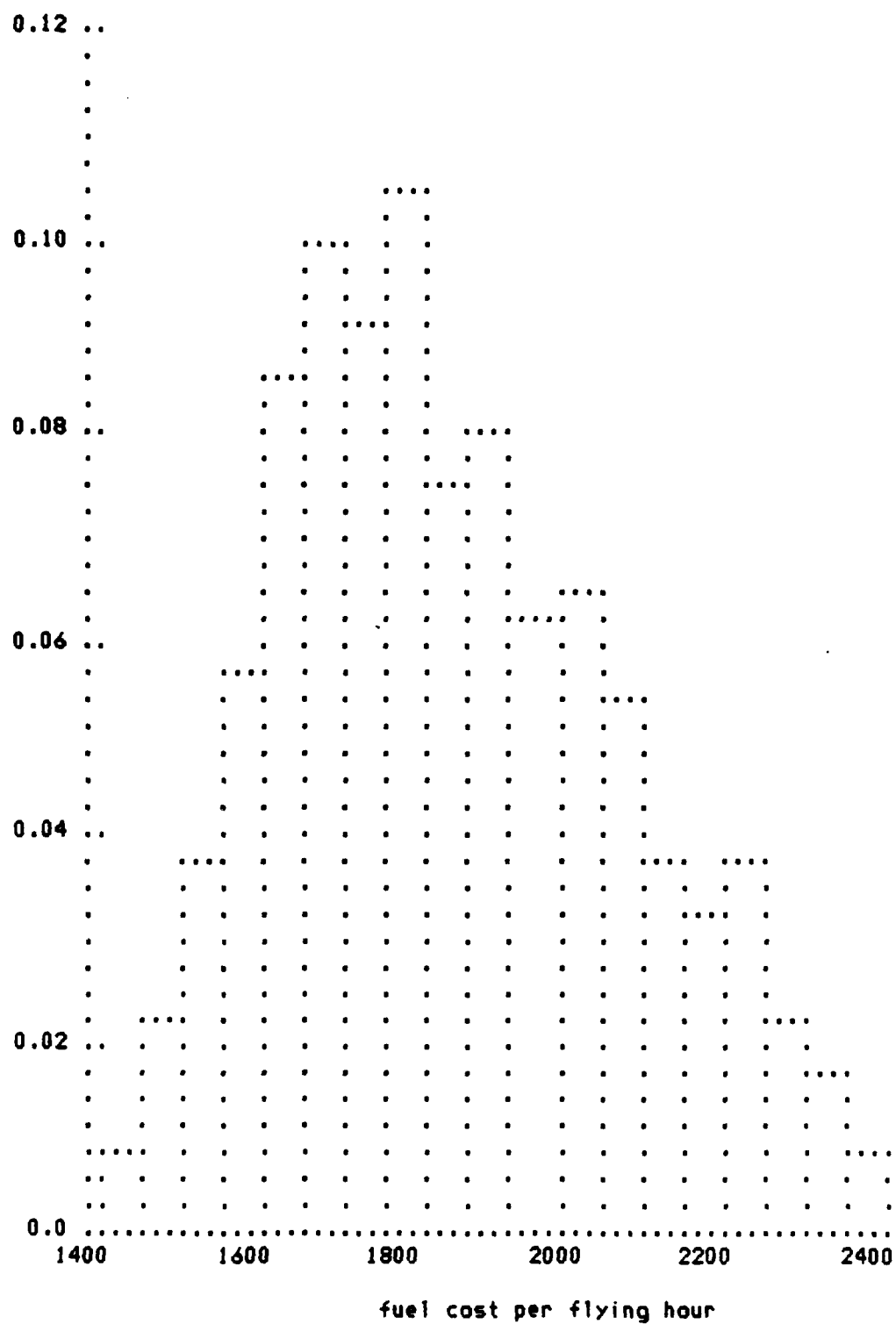
Substitution of the values for L, M, H, and y for each data point in each cost element yielded the values for the variable 'TX' in the Kolmogorov test (see line numbers 720-780 in Appendix H).

Appendix E: Fuel Cost Element Histograms

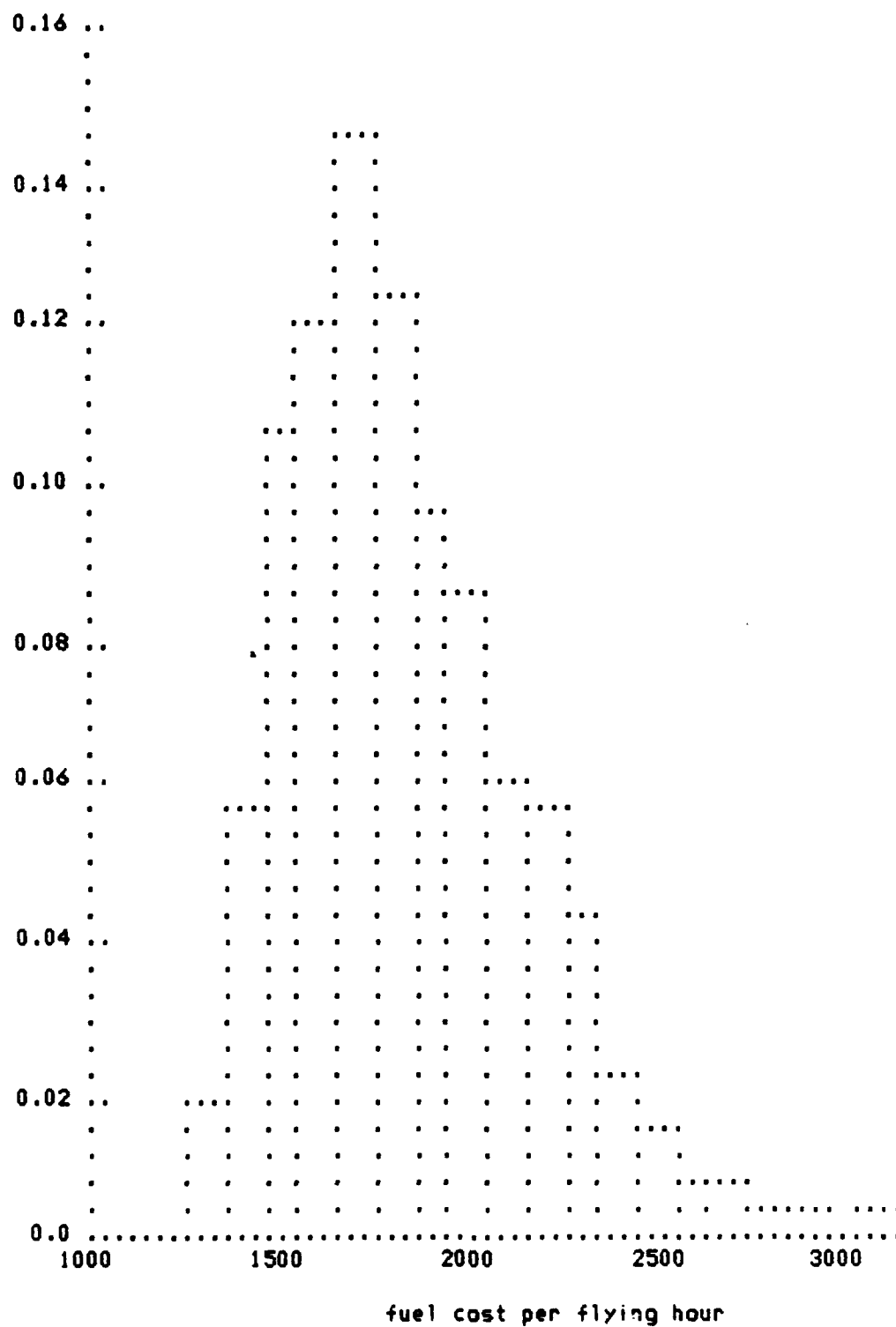
Baseline Data Relative Frequency Histogram



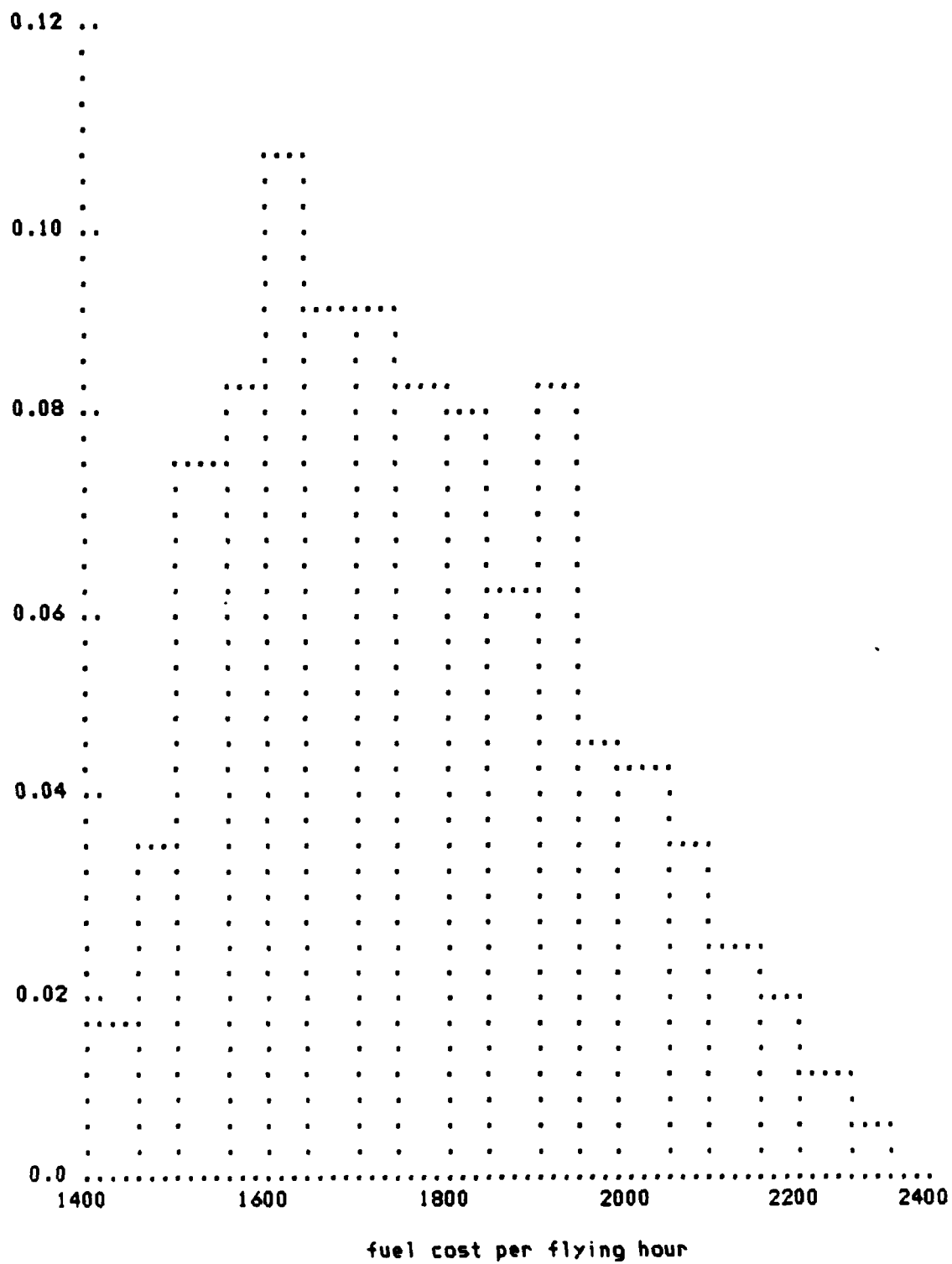
Fuel Sample Relative Frequency Histogram (Triangular Distribution)



Fuel Sample Relative Frequency Histogram (Gamma Distribution)

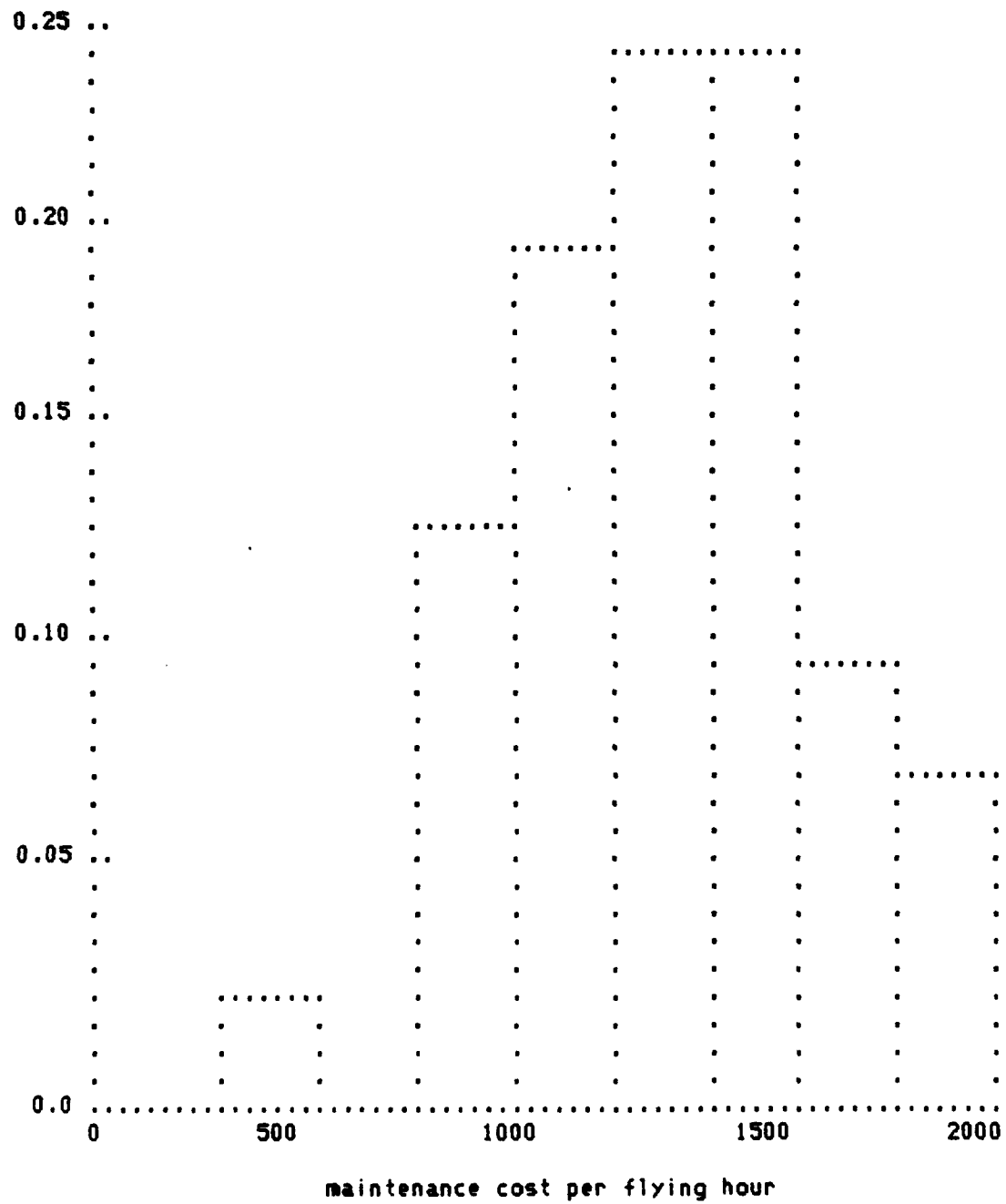


Fuel Sample Relative Frequency Histogram (Beta Distribution)

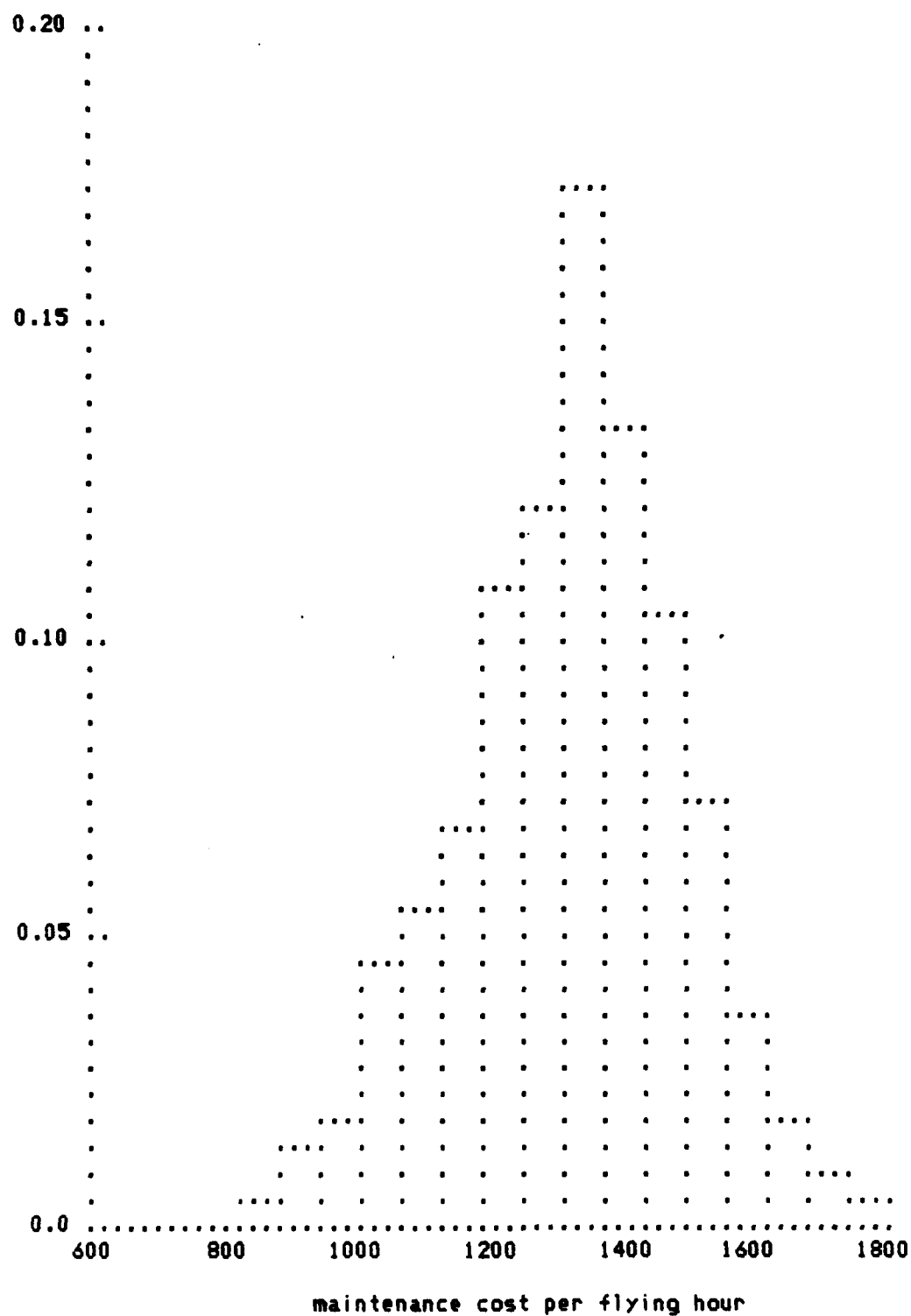


Appendix F: Maintenance Cost Element Histograms

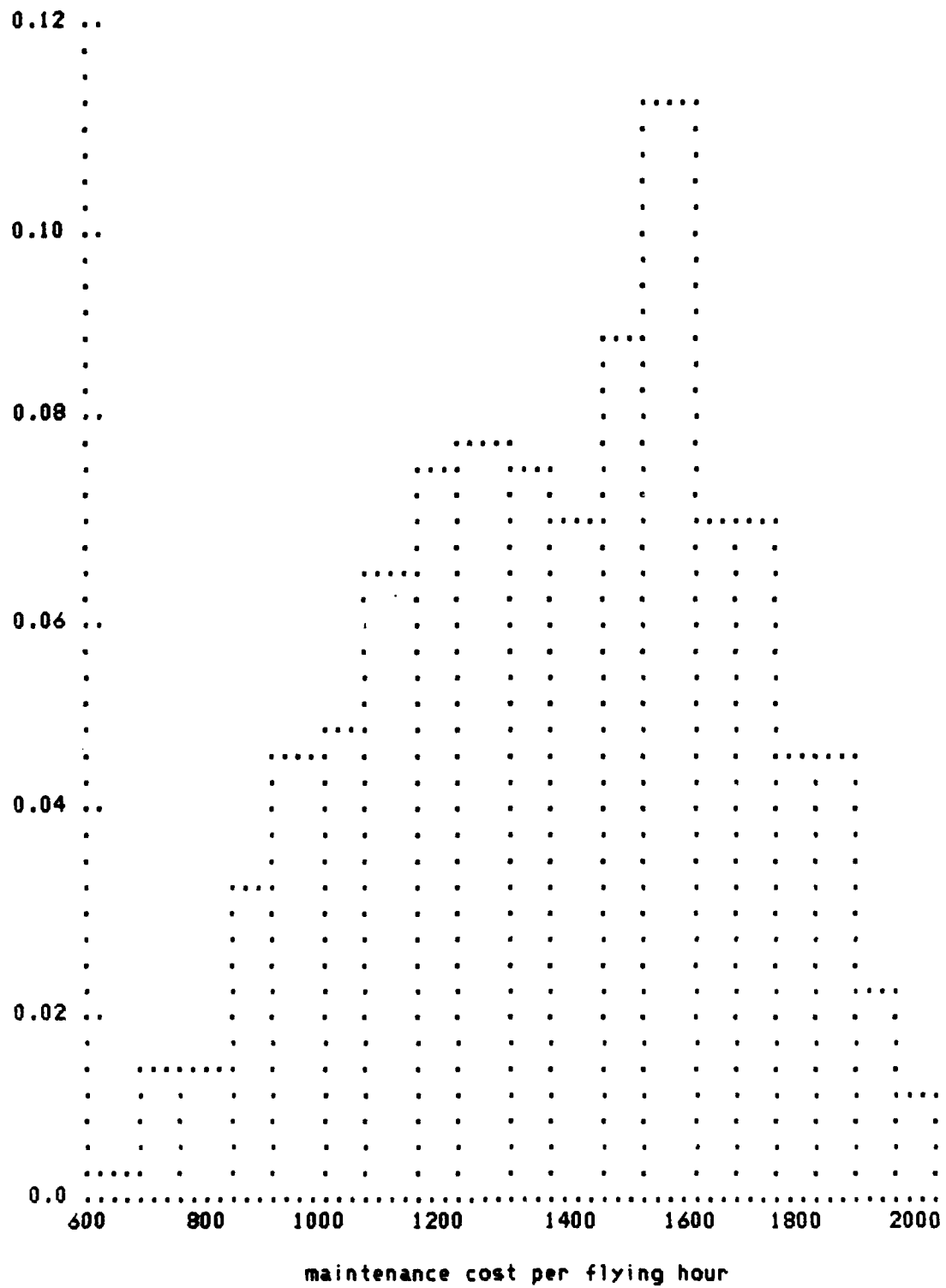
Baseline Data Relative Frequency Histogram



Maintenance Sample Relative Frequency Histogram (Triangular Distribution)

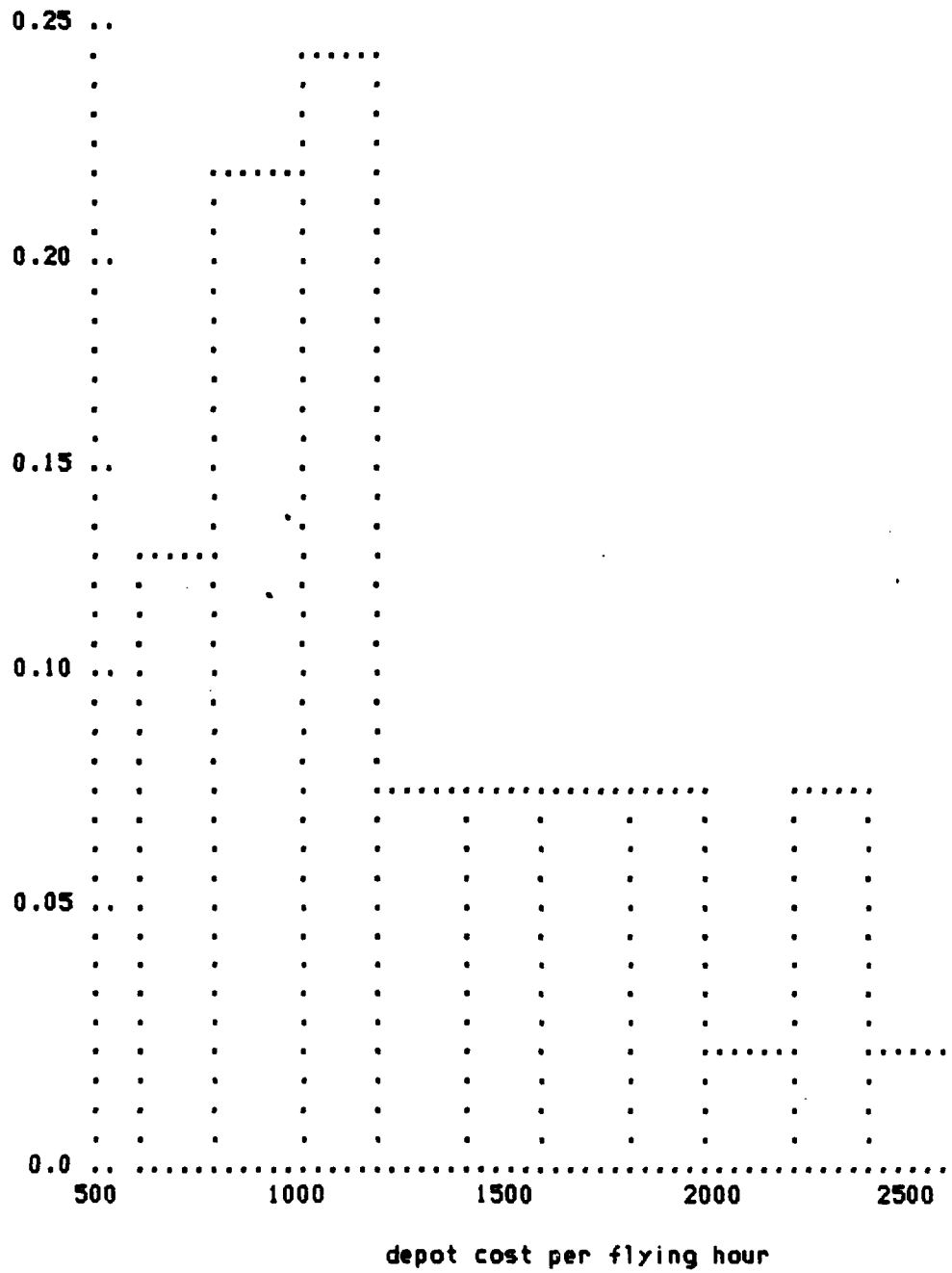


Maintenance Sample Relative Frequency Histogram (Beta Distribution)

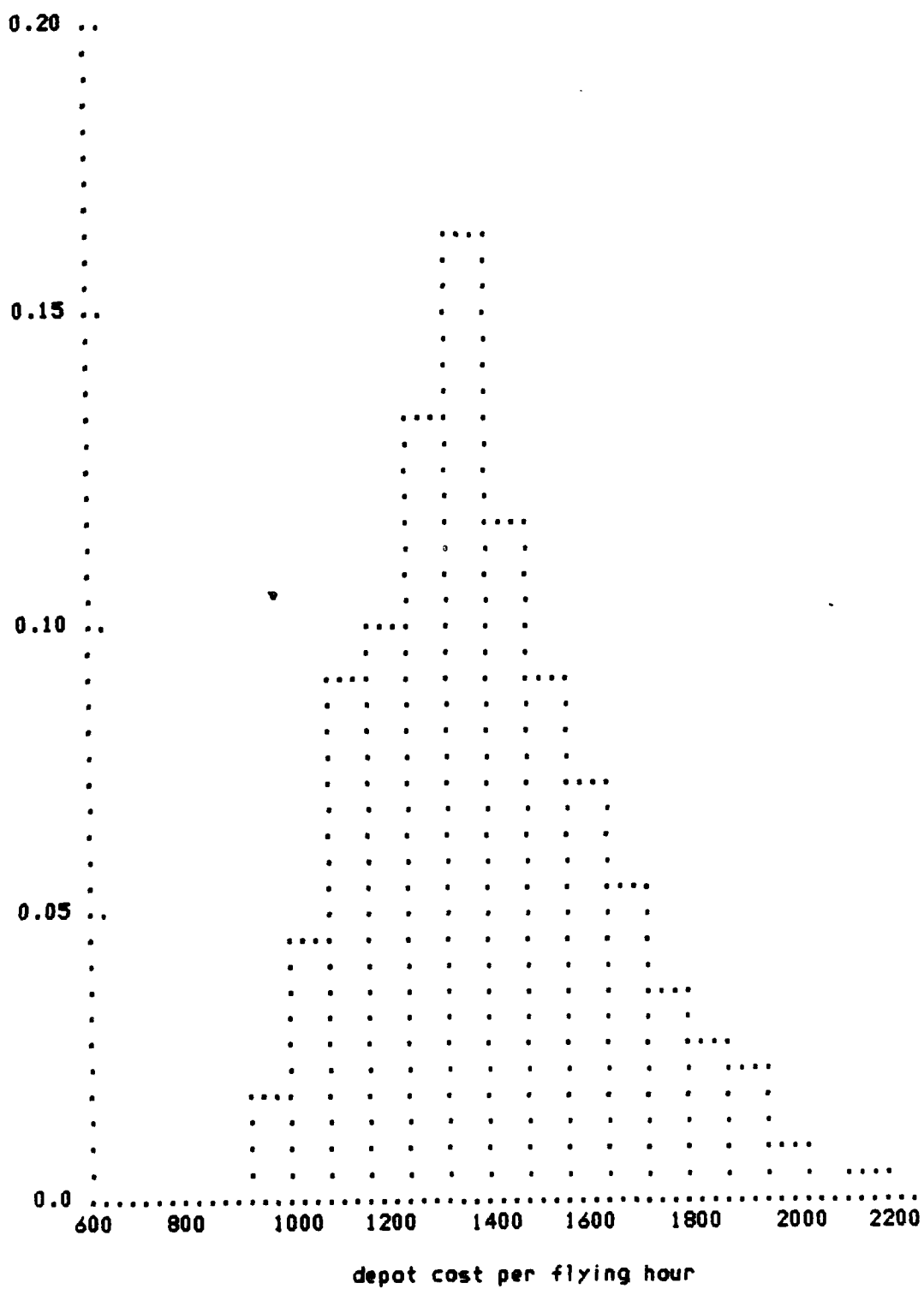


Appendix G: Depot Cost Element Histograms

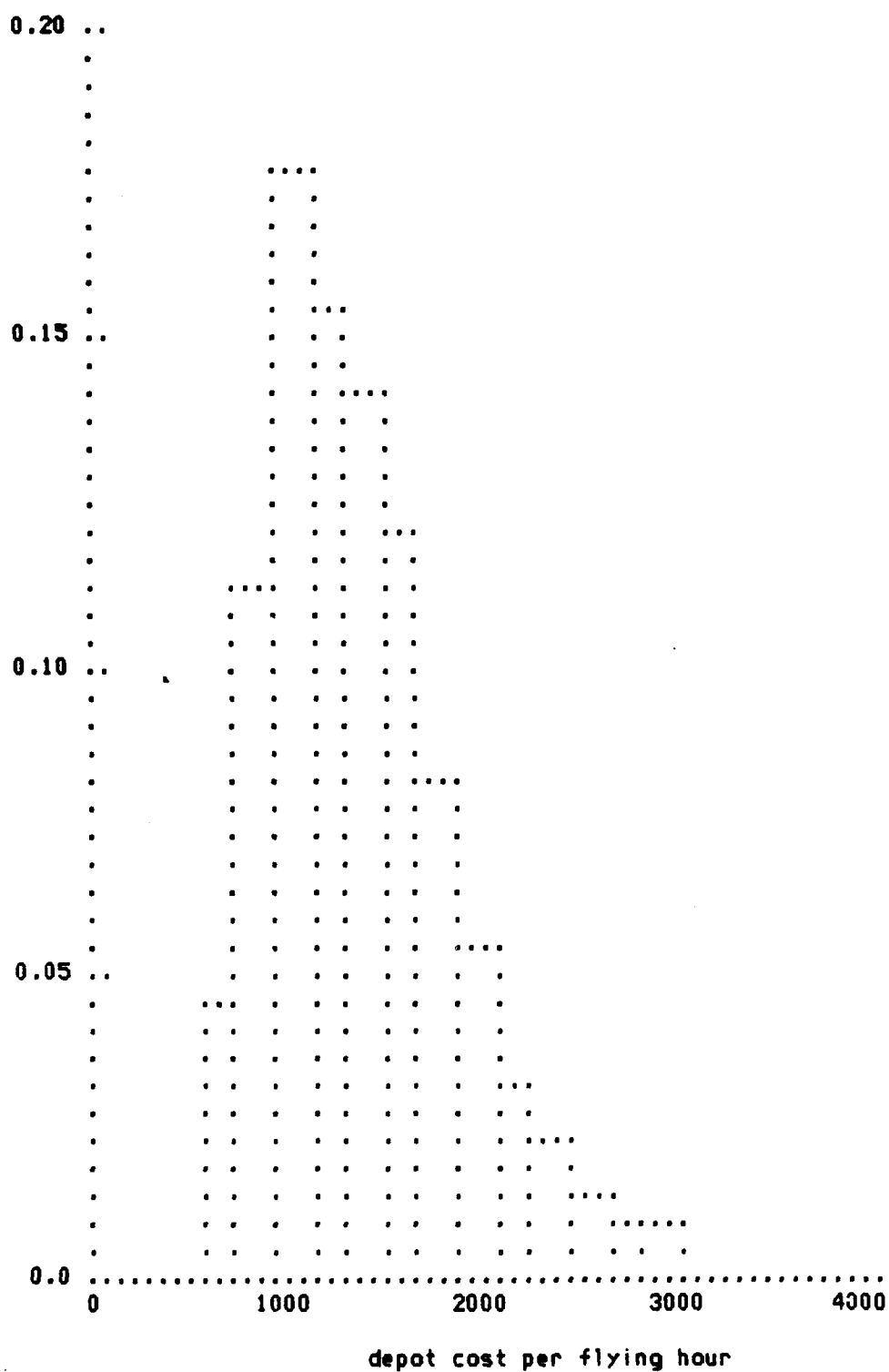
Baseline Data Relative Frequency Histogram



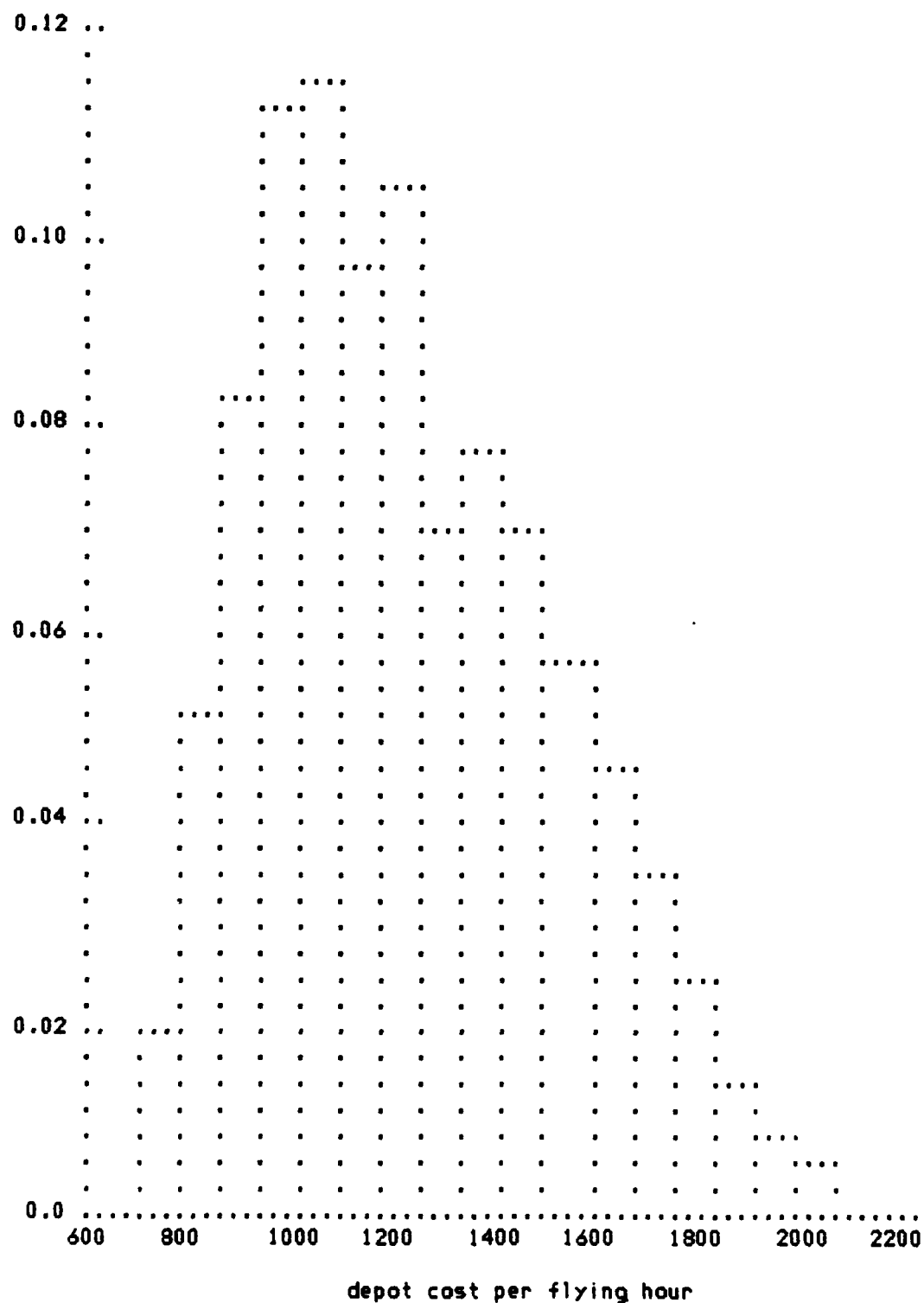
Depot Sample Relative Frequency Histogram (Triangular Distribution)



Depot Sample Relative Frequency Histogram (Gamma Distribution)



Depot Sample Relative Frequency Histogram (Beta Distribution)



Appendix H: Kolmogorov Test Source Code

```

100 REM KOLMOGOROV TEST - INTERACTIVE
110 REM
120 REM
130 REM      VARIABLE DEFINITIONS
140 REM
150 REM      X.....BASELINE DATA VALUES (ARRAY)
160 REM      XB,XG.....BET AND GAMMA DEVIATES (ARRAYS)
170 REM      SX.....PROBABILITY FOR EACH X (ARRAY)
180 REM      TX,GX,BX.....PROBABILITY FOR EACH X IN THE THEORETICAL
190 REM                      TRIANGULAR, GAMMA AND BETA DISTRIBUTIONS
200 REM      DT,DG,DB.....ABSOLUTE VALUE OF THE DIFFERENCE BETWEEN
210 REM                      THEORETICAL PROBABILITIES AND SX (ARRAYS)
220 REM      A,B,C.....GAMMA DISTRIBUTION PARAMETERS
230 REM      P,Q.....BETA DISTRIBUTION PARAMETERS
240 REM      U,PG.....ENTRY PARAMETERS FOR THE INCOMPLETE GAMMA
250 REM                      FUNCTION TABLES
260 REM      L,M,H.....LOW, MODAL, AND HIGH VALUES OF THE
270 REM                      BASELINE DATA SETS
280 REM      I.....LOOP COUNTER
290 REM      FLAG.....'1' MEANS CONTINUE SORT OR TEST FOR THE
300 REM                      GAMMA DISTRIBUTION
310 REM      YN.....FLAG TO CONTINUE PROGRAM FOR MORE DATA
320 REM      CE$.....COST ELEMENT NAME
330 REM
340 REM *****
350 REM
360 REM      READ IN AND SORT DATA
370 REM
380      DIM X(41),SX(41),TX(41),GX(41),BX(41)
390      DIM DT(41),DG(41),DB(41),XB(41),XG(41)
400      INPUT "WHAT COST ELEMENT";CE$
410      FOR I = 1 TO 41
420          INPUT "DATAPOINT IS";X(I)
430      NEXT I
440 REM
450 REM      SORT ROUTINE
460 REM
470      FLAG = 1
480      IF FLAG <> 1 GOTO 580
490          FLAG = 0
500          FOR I = 2 TO 41
510              IF X(I) > X(I - 1) THEN GOTO 560
520                  TEMP = X(I)
530                  X(I) = X(I - 1)
540                  X(I - 1) = TEMP
550              FLAG = 1
560          NEXT I

```

```

570          GOTO 480
580 REM      END SORT
590 REM
600 REM      *****
610 REM
620 REM      INPUT BASELINE PARAMETERS
630 REM
640          INPUT "WHAT ARE LOW, MODE AND HIGH (L,M,H)";L,M,H
650          INPUT "WHAT ARE GAMMA PARAMETERS (A,B,C)";A,B,C
660          INPUT "WHAT ARE BETA PARAMETERS (P,Q)";P,Q
670 REM
680 REM      *****
690 REM
700 REM      TRIANGULAR TEST
710 REM
720          FOR I = 1 TO 41
730              SX(I) = I/41
740              IF X(I) > M GOTO 770
750                  TX(I) = (X(I)-L)^2/((H-L)*(M-L))
760                  GOTO 780
770                  TX(I) = 1 - ((B-X(I))^2/((H-L)*(H-M)))
780                  DT(I) = ABS(TX(I)-SX(I))
790              NEXT I
800 REM
810          LPRINT CE$
820          FOR I = 1 TO 41
830              LPRINT X(I),SX(I),TX(I),DT(I)
840          NEXT I
850          LPRINT
860 REM
870 REM      *****
880 REM
890 REM      BETA TEST
900 REM
910          FOR I = 1 TO 41
920              XB(I) = (X(I)-L)/(H-L)
930              PRINT XB(I)
940              INPUT "WHAT IS THE VALUE FROM THE GRAPH";BX(I)
950              DB(I) = ABS(BX(I)-SX(I))
960              LPRINT XB(I),SX(I),BX(I),DB(I)
970          NEXT I
980          LPRINT
990 REM
1000 REM      *****
1010 REM
1020 REM      GAMMA TEST
1030 REM
1040          FLAG = 0
1050          INPUT " IS GAMMA TEST NEEDED? (1=YES)",FLAG
1060          IF FLAG (<>) 1 GOTO 1180
1070          FOR I = 1 TO 41
1080              XG(I) = (X(I)-C)/B
1090              U = XG(I)/SQR(A)

```

```

1100      PG = A-1
1110      PRINT U,PG
1120      INPUT "ENTER TABLE VALUE",GX(I)
1130      DG(I) = ABS(GX(I)-SX(I))
1140      LPRINT XG(I),SX(I),GX(I),DG(I)
1150      NEXT I
1160      LPRINT
1170 REM
1180      INPUT "ANALYZE ANOTHER COST ELEMENT? (1=YES)",YN
1190      IF YN = 1 GOTO 400
1200 REM
1210      END
1220 REM *****

```

Bibliography

1. Aircraft Operating and Support Cost Development Guide. Washington: Office of the Secretary of Defense, Cost Analysis Improvement Group, 15 April 1980.
2. Bell, Archie C. and Capt Daniel P. Turney. A Summary and Analysis of the Initial Application of Life Cycle Costing Techniques to a Major Weapon System Acquisition, MS thesis, LSSR 35-78B. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1978 (AD-A061 304).
3. Blanchard, Benjamin S. Logistics Engineering and Management. Englewood Cliffs NJ: Prentice-Hall Inc., 1981.
4. Caver, Lt Col Troy L. "Life Cycle Cost: Attitudes and Latitudes," Defense Management Journal, 15: 12-17 (July 1979).
5. Cira, Capt Anthony T. and Capt Kenneth R. Jennings. Life Cycle Costing: A Working Level Approach, MS thesis LSSR 37-81, School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, June 1981 (AD-A103 256).
6. Conover, W.J. Practical Nonparametric Statistics. New York: John Wiley and Sons, Inc., 1971.
7. Coon, Helen. "Note on William A. Donaldson's 'The Estimation of the Mean and Variance of a 'PERT' Activity Time'," Operations Research, 13: 386-387 (May-June 1965).
8. Davis, Capt William R. and Capt John R. Wysowski. A Summary and Analysis of the Logistics Support Cost Model Application to the ACF/F-16 Weapon System Acquisition, MS thesis, LSSR 6-79A. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, June 1979 (AD-A072 592).
9. Department of Defense. Life Cycle Costing Guide for Systems Acquisition (Interim). Washington: Government Printing Office, January 1973.
10. Department of Defense. Major Systems Acquisition Procedures. DoD Instruction 5000.2. Washington: Government Printing Office, 8 March 1983.
11. Department of the Air Force. USAF Cost and Planning Factors. AFR 173-13. Washington: HQ USAF, 1 February 1984.

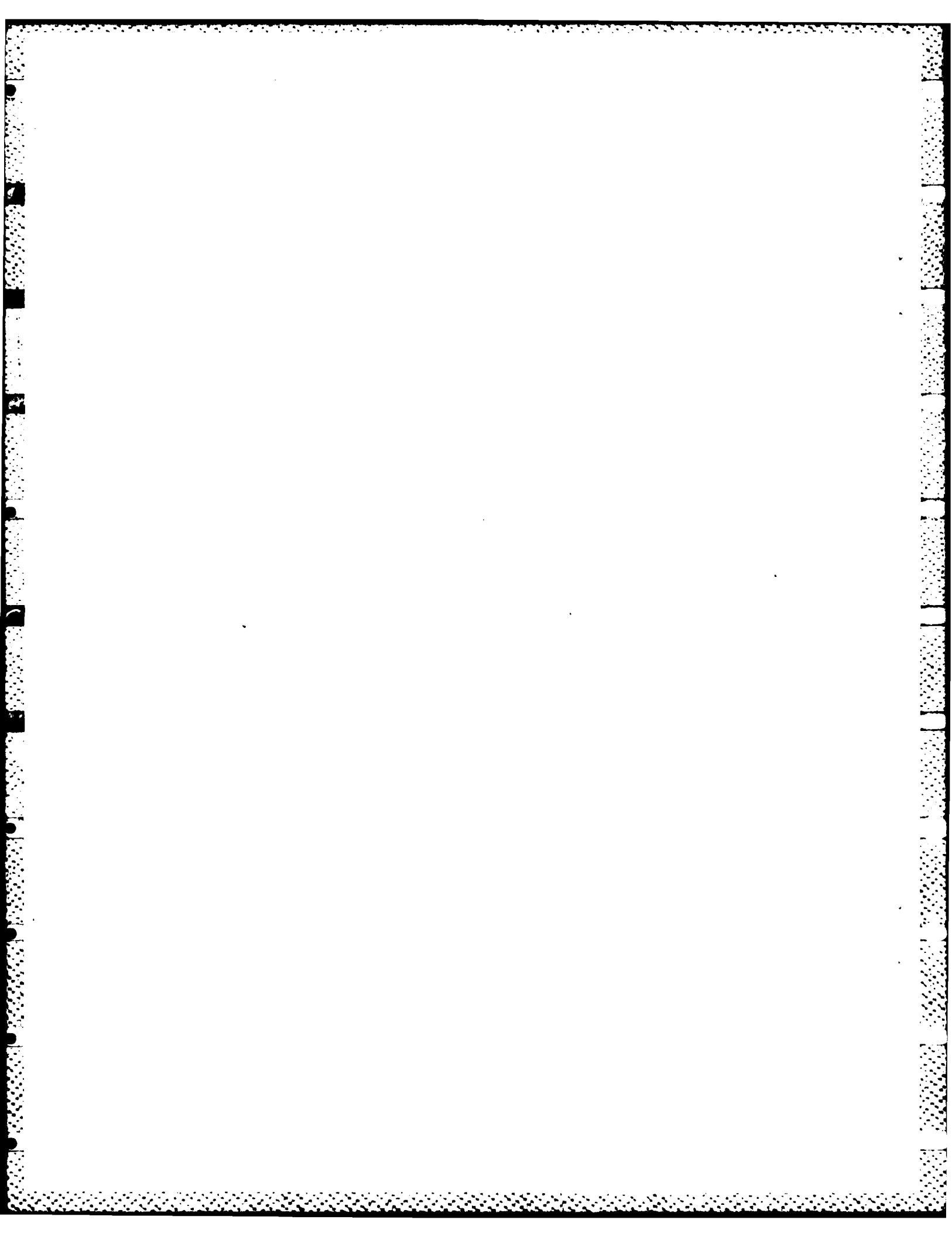
12. Department of the Air Force. Life Cycle Cost Management Program. AFR 800-11. Washington: HQ USAF, 22 February 1978.
13. Dervitsiotis, Kostas N. Operations Management. New York: McGraw-Hill Inc., 1981.
14. Dienemann, Paul F. Estimating Cost Uncertainty Using Monte Carlo Techniques. RM-4854-PR. Santa Monica CA: The Rand Corporation, January 1966 (AD629 082).
15. Donaldson, William A. "The Estimation of the Mean and Variance of a 'PERT' Activity Time." Operations Research, 13: 382-385 (May-June 1965).
16. Emmelhainz, Margaret A. "Innovative Contractual Approaches to Controlling Life Cycle Costs," Defense Management Journal, 19: 36-42 (2nd Quarter 1983).
17. Executive Office of the President, Office of Management and Budget. The United States Budget in Brief, Fiscal Year 1984. Washington: Government Printing Office, 1983.
18. Fiorello, Marco R. Problems in Avionics Life-Cycle Analysis. P-5136. Santa Monica CA: The Rand Corporation, December 1973 (AD783 320).
19. Fisher, G.H. Cost Functions and Budgets. P-3789. Santa Monica CA: The Rand Corporation, February 1968.
20. -----, The Analytical Bases of Systems Analysis. P-3363. Santa Monica CA: The Rand Corporation, May 1966 (AD634 512).
21. Fisher, Irving N. and George R. Hall. Risk and the Aerospace Rate of Return. RM-5440-PR. Santa Monica CA: The Rand Corporation, December 1967 (AD643 726).
22. Giffin, Walter C. Introduction To Operations Engineering. Homewood IL: Richard D. Irwin, Inc., 1971.
23. Harter, H. Leon. New Tables of the Incomplete Gamma-Function Ratio and of Percentage Points of the Chi-Square and Beta Distributions. Washington: Government Printing Office, 1964.
24. Holloway, Charles A. Decision Making Under Uncertainty. Englewood Cliffs NJ: Prentice Hall, Inc., 1979.
25. Kankey, Roland D. "Precepts for Life Cycle Cost Management," Air Force Journal of Logistics, 6: 28-30 (Winter 1982).
26. Kaufmann, William W. "The Defense Budget," Setting National Priorities, the 1984 Budget. Edited by Joseph A. Pechman. Washington: The Brookings Institute, 1983.

27. Klipfel, Stephen R. "The Independent Cost Analysis Program at AFLC," The Air Force Comptroller, 13: 30-31 (October 1979).
28. Lib-0008. IMSL Library. Reference Manual. IMSL, Inc., Houston TX, June 1980.
29. Long, John Amos. Life Cycle Costing In a Dynamic Environment. PhD dissertation. Columbus OH: Ohio State University, 1983.
30. Marks, K.E. and H. G. Massey. Life Cycle Analysis Procedures and Techniques: An Appraisal and Suggestions for Future Research. P-6031. Santa Monica CA: Rand Corporation, October 1977.
31. Marsh, General Robert T. "Project Cost: A New Approach," TIG Brief, 35: 1-2 (21 February 1983).
32. Moore, Peter G. Risk In Business Decision. New York: John Wiley and Sons, Inc., 1972.
33. Naylor, Thomas H. and others. Computer Simulation Techniques. New York: John Wiley and Sons, Inc., 1966.
34. Office of Management and Budget. "Major Systems Acquisition," Circular No. A-109. Washington: Government Printing Office, April 5, 1976.
35. Perry, C. and I. D. Grieg. "Estimating the Mean and Variance of Subjective Distributions in PERT and Decision Analysis." Management Science, 21: 1477-1480 (August 1975).
36. Recktenwalt, Thomas J. "Visibility and Management of Operating and Support Costs, System II (VAMOSC II)." Presentation to the Sixteenth Annual DoD Cost Analysis Symposium, Arlington VA, October 1981 (AD-A109 308).
37. Seldon, M. Robert. Life Cycle Costing: A Better Method of Government Procurement. Boulder CO: Westview Press, Inc., 1979.
38. Singer, Robert A. "Acquiring Affordable Weapon Systems." Study Project Report 77-1. Defense Systems Management College, Ft. Belvoir VA, May 1977 (AD-A042 777).
39. Worm, George H. Applied Risk Analysis With Dependence Among Cost Components. Contract F33615-81-C-5138. Clemson University, Clemson SC, November 1981.

VITA

Major David B. Freeman was born on 7 April 1949 in Kalamazoo, Michigan. He graduated from high school in Johnstown, Pennsylvania, in 1967 and went on to attend Juniata College in Huntingdon, Pennsylvania. He received the degree of Bachelor of Science in Biology in June 1971. In September 1972, he received his commission in the USAF through OTS. He completed navigator training and received his wings in August 1973. He served as an F-4E Weapon Systems Officer at Udorn RTAFB, Thailand, Eglin AFB, Florida, and Clark AB, Philippines. He was then assigned to the 479 TFW, Holloman AFB, New Mexico, where he served as an AT-388 flight and academic instructor until entering the School of Systems and Logistics, Air Force Institute of Technology in June 1983.

Permanent Address: 5909 Hartwick Lane
Dayton, Ohio 45424



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFIT/GLM/LSM/84S-24			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION School of Systems and Logistics		6b. OFFICE SYMBOL (If applicable) AFIT/LS		7a. NAME OF MONITORING ORGANIZATION	
6c. ADDRESS (City, State and ZIP Code) Air Force Institute of Technology Wright-Patterson AFB, Ohio 45433				7b. ADDRESS (City, State and ZIP Code)	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION		8b. OFFICE SYMBOL (If applicable)		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. ADDRESS (City, State and ZIP Code)		10. SOURCE OF FUNDING NOS.			
		PROGRAM ELEMENT NO.		PROJECT NO.	TASK NO.
11. TITLE (Include Security Classification) See Box 19		WORK UNIT NO.			
12. PERSONAL AUTHOR(S) David B. Freeman, B.S., Major, USAF					
13a. TYPE OF REPORT MS Thesis		13b. TIME COVERED FROM _____ TO _____		14. DATE OF REPORT (Yr., Mo., Day) 1984 September	
				15. PAGE COUNT 92	
16. SUPPLEMENTARY NOTATION <div style="text-align: right;">Approved for public release LAW APR 1984 John E. Wolaver Dean for Research and Professional Development</div>					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB. GR.			
15	05		Logistics, Costs, Cost Estimates, VAMOSC,		
12	02		Life Cycle Costs, Risk, Military Procurement		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) Title: THE EFFECT OF ASSUMPTIONS ABOUT COST ELEMENT PROBABILITY DISTRIBUTIONS ON OPERATING AND SUPPORT COST RISK ANALYSIS Thesis Chairman: John A. Long, Lieutenant Colonel, USAF					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>			21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED		
22a. NAME OF RESPONSIBLE INDIVIDUAL John A. Long, Lt Col, USAF			22b. TELEPHONE NUMBER (Include Area Code) 513-255-4845		22c. OFFICE SYMBOL AFIT/LSY

The purpose of this study was to determine the effects of assumptions about cost element probability distributions on Operating and Support cost estimates. Currently, these assumptions are being made arbitrarily, without regard for their effect on the risk associated with a total cost estimate.

The experiment was performed with a Monte Carlo simulation. Three cost elements, fuel costs, maintenance personnel costs, and depot maintenance costs were the sample space. Historical data from 41 fighter aircraft was used to determine the low, high, and modal values for each cost element. The Triangular, Beta, and Gamma distributions were selected as candidate distributions for the cost elements. The low, modal, and high values provided a means by which to parameterize the distributions for specific cost elements. The distribution for each cost element was varied over the set of candidate distributions with all other factors held constant, samples were drawn, and the samples summed to provide total cost estimates. The result was an array of 18 total cost distributions containing 1000 data points each.

The variance of the 18 resultant distributions was the focus of the analysis. That analysis indicated that the cost element distributions chosen do have an effect on the risk of the total cost estimate. Further, distributions with infinite upper bounds result in consistently higher risk than those with finite upper bounds.

END

FILMED

2-84

DTIC